

Université de Montréal

Sur les prolongements de sous-copules

par

Ayi Ajavon

Département de mathématiques et de statistique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Statistique

février 2015

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Sur les prolongements de sous-copules

présentée par

Ayi Ajavon

a été évaluée par un jury composé des personnes suivantes :

Professeur Alejandro Murua

(président-rapporteur)

Professeur François Perron

(directeur de recherche)

Professeur Manuel Morales

(membre du jury)

Professeur Ernest Monga

(examineur externe)

Professeur Patrice Marcotte

(représentant du doyen de la FES)

Thèse acceptée le :

16 janvier 2015

RÉSUMÉ

L'objet du travail est d'étudier les prolongements de sous-copules. Un cas important de l'utilisation de tels prolongements est l'estimation non paramétrique d'une copule par le lissage d'une sous-copule (la copule empirique). Lorsque l'estimateur obtenu est une copule, cet estimateur est un prolongement de la sous-copule. La thèse présente au chapitre 2 la construction et la convergence uniforme d'un estimateur bona fide d'une copule ou d'une densité de copule. Cet estimateur est un prolongement de type copule empirique basé sur le lissage par le produit tensoriel de fonctions de répartition splines. Le chapitre 3 donne la caractérisation de l'ensemble des prolongements possibles d'une sous-copule. Ce sujet a été traité par le passé par [Schweizer et Sklar, 1974], [Nešlehová, 2007], [Rüschendorf, 2009], [Mesfioui et Quessy, 2010], [de Amo, Díaz Carrillo et Fernández-Sánchez, 2012] ; mais les constructions proposées ne s'appliquent pas à la dépendance dans des espaces très généraux. Le chapitre 4 s'attèle à résoudre le problème suivant posé par [Carley, 2002]. Il s'agit de trouver la borne supérieure des prolongements en dimension 3 d'une sous-copule de domaine fini.

Sommaire et mots-clés français : copules, sous-copules, densités, prolongements, borne supérieure.

SUMMARY

The extension of subcopulas is an important domain. One of possible applications is the nonparametric estimation of a copula: it consists of the smoothing of a subcopula (the empirical copula) while preserving the copulas properties.

In Chapter 2, we present an extension of the empirical copula based on the tensor product of splines functions. Our estimators are bona fide estimators of the copula.

Chapter 3 tackles the problem of finding all possible extensions of a given subcopula. This subject has been treated by [Schweizer et Sklar, 1974], [Nešlehová, 2007], [Rüschendorf, 2009], [Mesfioui et Quessy, 2010], [de Amo et coll., 2012]; but these characterizations do not apply on very general spaces.

Chapter 4 deals with the following problem of [Carley, 2002]: finding the expression of the upper bound of the extensions of a finite subcopula in dimension 3.

English summary and keywords: copulas, subcopulas, density, extensions, upper bound.

Table des matières

Résumé.....	v
Summary	vii
Remerciements	1
Chapitre 1. Introduction	3
1.1. Les copules et la transformation de Feller [Feller, 1968]	4
1.2. Les différents types de prolongements de sous-copule	6
1.2.1. Les prolongements analytiques	7
1.2.1.1. Le prolongement bilinéaire	7
1.2.1.2. Les prolongements basés sur la méthode des rectangles	9
1.2.2. Les prolongements probabilistes	10
1.2.3. Le cas particulier de la copule empirique et de la copule empirique lissée	10
1.3. Bornes de Fréchet des prolongements	12
1.3.1. Bornes de Fréchet des prolongements en dimension 2	12
1.3.2. Bornes de Fréchet des prolongements en dimension $d > 2$	12
1.4. Apports de la thèse et conclusion	13
Chapitre 2. Convergence des estimateurs bona fide de copule et de densité de copule obtenus par lissage spline de la copule empirique	15
2.1. Introduction	15
2.2. Approximation de copules par des splines	17
2.3. Estimation de la copule et de la densité de copule : principaux résultats	21
2.3.1. Estimation de la copule	21
2.3.2. Estimation de la densité de copule	25

2.4.	Discussion	31
2.5.	Résultats intermédiaires et preuves des théorèmes	31
2.6.	Conclusion	34
Chapitre 3. Caractérisation des prolongements d'une sous-copule en dimension 2		43
3.1.	Introduction	43
3.2.	Les prolongements	45
3.3.	La copule maximale	52
3.4.	La copule minimale	55
3.5.	La méthode des rectangles et l'ensemble de Cantor	58
3.6.	Conclusion	61
Chapitre 4. Borne supérieure des prolongements d'une sous-copule à domaine fini en dimension 3		63
4.1.	Introduction	63
4.2.	Caractérisation des prolongements d'une sous-copule avec un domaine de cardinalité finie	65
4.3.	La solution en dimension 2	69
4.4.	La borne supérieure en dimension 3	70
4.5.	Conclusion	85
Chapitre 5. Conclusion		87
Annexe A. Une approche probabiliste des splines		A-i
A.1.	Construction de splines monotones	A-i
A.2.	Approximation à l'aide des splines monotones	A-v
Annexe B. Démonstration du lemme 3.2.2		B-i
Bibliographie		B-i

REMERCIEMENTS

Mes sentiments de gratitude vont à mon superviseur, François Perron, qui m'a offert l'assistance et la direction pour l'accomplissement de ce travail. Son aide pour la préparation de l'examen de synthèse m'a été précieuse. Avoir été son démonstrateur aux cours de probabilité et de processus stochastiques m'a permis d'approfondir mes connaissances en mathématique discrète et en processus stochastiques.

Je remercie aussi le Département de mathématiques et statistique de l'Université de Montréal qui m'a accueilli durant toutes ces années de thèse et m'a octroyé une bourse d'étude pour l'accomplissement de ce travail. J'ai trouvé un corps professoral disponible et rigoureux. Je veux remercier ici : le Professeur Christian Léger du cours de Consultation Statistique pour sa disponibilité et sa rigueur ; le professeur Alejandro Murua du cours de méthodes asymptotiques pour m'avoir introduit à l'asymptotique des estimateurs de maximum de vraisemblance ; la professeure Mylène Bédard pour avoir suscité mon intérêt pour la statistique bayésienne. Je ne saurais oublier le professeur Martin Bilodeau dont la rigueur d'exposition du matériel de cours est éblouissant et le professeur Sabin Lessard du cours de probabilité à qui je dois une plus grande maîtrise de la convergence presque sûre.

Mes remerciements vont aussi à la Faculté des Études Supérieures de l'Université de Montréal qui m'a octroyé une bourse de quatrième année pour finir la thèse, à mes collègues du Département de mathématiques et statistique, Folly, Blache, Valéry, Hervé, Joseph, Joel, qui de près ou de loin ont contribué à la réalisation de ce travail.

Chapitre 1

INTRODUCTION

Ce chapitre introductif présente les notions de copules et sous-copules et donne la motivation du sujet de recherche qu'est le prolongement des sous-copules. Il donne une vue générale des résultats de la thèse.

La dépendance entre des variables aléatoires réelles X_1, X_2, \dots, X_d , est décrite par la donnée de la fonction de répartition conjointe

$$F : (-\infty, +\infty)^d \mapsto [0, 1]$$
$$F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d),$$

$x_1, \dots, x_d \in \mathbb{R}$. Rappelons que les fonctions de répartition marginales associées sont données par

$$F_i(x_i) = P(X_i \leq x_i),$$

$x_i \in \mathbb{R}, i = 1, \dots, d$. L'étude des copules découle du problème de Fréchet [Fréchet, 1956] consistant à trouver l'ensemble des fonctions de répartition conjointes F compatibles avec des fonctions de répartition marginales fixées, F_1, \dots, F_d . En dimension $d = 2$, il a montré que les bornes de cet ensemble sont

$$W(F_1(x_1), F_2(x_2)) \leq F(x_1, x_2) \leq M(F_1(x_1), F_2(x_2)), \quad x_1, x_2 \in \mathbb{R},$$
$$W(u_1, u_2) = \max(u_1 + u_2 - 1, 0), \quad (u_1, u_2) \in [0, 1]^2,$$
$$M(u_1, u_2) = \min(u_1, u_2), \quad (u_1, u_2) \in [0, 1]^2.$$

Les fonctions $W(F_1(\cdot), F_2(\cdot))$, $M(F_1(\cdot), F_2(\cdot))$, sont des fonctions de répartition conjointes appelées bornes de Fréchet. Wassilly Hoeffding [Hoeffding, 1940] fut le premier à ramener le problème à celui de la recherche de l'ensemble des fonctions de répartition conjointes F compatibles avec des fonctions de répartition marginales uniformes sur $[-1/2, 1/2]$. Il anticipait par là l'introduction de la notion des copules par Sklar [Sklar, 1959]. En effet, les copules seront définies comme des fonctions de répartition sur $[0, 1]^2$ dont les marges sont uniformes (lorsque les fonctions de répartition marginales sont continues).

En toute généralité, une d -copule est une fonction

$$C : [0, 1]^d \rightarrow [0, 1]$$

qui satisfait les critères suivants :

- $C(u_1, \dots, u_d) = u_i$, si $u_1 u_2 \cdots u_d = u_i$, $i = 1, \dots, d$;
- C est d -croissante c'est-à-dire pour $\mathbf{u} = (u_1, \dots, u_d)$, $\mathbf{v} = (v_1, \dots, v_d)$ tels que $u_1 \leq v_1, \dots, u_d \leq v_d$, le C -volume du pavé $[\mathbf{u}, \mathbf{v}] = \mathbf{X}_{i=1}^d [u_i, v_i]$, $V_C([\mathbf{u}, \mathbf{v}])$ est positif ou nul, et est donné par

$$V_C([\mathbf{u}, \mathbf{v}]) = \sum_{t \in \mathbf{X}_{i=1}^d \{u_i, v_i\}} (-1)^{n(t)} C(t),$$

où $n(t) = \#\{k : t_k = u_k, k = 1, \dots, d\}$, $t \in \mathbf{X}_{i=1}^d \{u_i, v_i\}$, \mathbf{X} étant mis pour le produit cartésien généralisé.

Une autre définition des copules est celle donnée par Nelsen [Nelsen, 1999, chapitre 2] qui définit les copules comme des fonctions reliant les distributions multivariées à leurs distributions marginales univariées. En effet, à tout point $(F_1(x_1), \dots, F_d(x_d)) \in [0, 1]^d$, on peut associer le scalaire $F(x_1, \dots, x_d)$, $x_i \in \mathbb{R}$, $i = 1, \dots, d$. Un résultat fondamental de Sklar [Sklar, 1959] énonce que cette correspondance est réalisée par des fonctions (appelées copules).

Théorème 1.0.1. *Étant donnée F une fonction de répartition sur \mathbb{R}^d avec F_1, \dots, F_d comme fonctions de répartition marginales, il existe une d -copule telle que pour $(x_1, \dots, x_d) \in \mathbb{R}^d$,*

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

C étant définie de manière unique sur $R_1 \times \cdots \times R_d$, R_i étant l'ensemble image de F_i , $i = 1, \dots, d$.

Les copules sont très utiles pour caractériser la dépendance entre des variables aléatoires dont les fonctions de répartition marginales sont connues.

1.1. LES COPULES ET LA TRANSFORMATION DE FELLER [FELLER, 1968]

Il est très fréquent dans l'étude des copules, de ramener les fonctions de répartition marginales à l'uniforme. Les transformations qui permettent une telle opération sont celles de Feller [Feller, 1968] et ses généralisations.

Soient X_1, X_2, \dots, X_d des variables aléatoires de fonctions de répartition marginales F_1, F_2, \dots, F_d continues et de fonction de répartition conjointe F continue. D'après le théorème de Sklar 1.0.1, il existe une copule C telle que $F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d))$, $(x_1, \dots, x_d) \in \mathbb{R}^d$. Considérons les variables aléatoires $U_1 =$

$F_1(X_1), U_2 = F_2(X_2), \dots, U_d = F_d(X_d)$ introduites par [Feller, 1968]. La fonction de répartition conjointe du vecteur (U_1, U_2, \dots, U_d) est la copule C , et les fonctions de répartition marginales de U_1, U_2, \dots, U_d sont uniformes. La transformation $U_i = F_i(X_i)$, $i = 1, \dots, d$ permet de ramener les fonctions de répartition marginales à la fonction de répartition de la loi uniforme sur $[0, 1]$ tout en conservant C comme copule caractérisant la dépendance. Cependant cette transformation n'est pas adaptée au cas où les variables aléatoires X_1, X_2, \dots, X_d ne sont pas continues. Voici une généralisation proposée par Szczesny [Szczesny, 1991].

Définition 1.1.1 ([Szczesny, 1991]). *Soit X une variable aléatoire de fonction de répartition F ; définissons la transformation I_F^**

$$I_F^*(F(x), u) = \begin{cases} 1, & \text{si } F(x) \leq u; \\ \frac{u - F(x)^-}{F(x) - F(x)^-}, & \text{si } F(x)^- \leq u < F(x); \\ 0, & \text{si } F(x)^- > u. \end{cases}$$

$x \in \mathbb{R}$, $F(x)^- = \sup\{y, y = F(t), t < x\}$.

Soient X_1, X_2, \dots, X_d des variables aléatoires continues de fonctions de répartition marginales F_1, F_2, \dots, F_d et de fonction de répartition conjointe F . Soit la fonction de répartition F^* définie par

$$F^*(u_1, \dots, u_d) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} I_{F_1}^*(F_1(x_1), u_1) \dots I_{F_d}^*(F_d(x_d), u_d) dF(x_1, \dots, x_d),$$

$(u_1, \dots, u_d) \in [0, 1]^d$. Aux points de continuité de F_1, F_2, \dots, F_d , nous avons la relation

$$F^*(F_1(x_1), \dots, F_d(x_d)) = F(x_1, \dots, x_d),$$

$(x_1, \dots, x_d) \in \mathbb{R}^d$.

Théorème 1.1.1 (Généralisation du théorème de Sklar, [Niewiadomska-Bugaj et Kowalczyk, 2005]). *Etant donnée F une fonction de répartition sur $\overline{\mathbb{R}}^d$ avec F_1, \dots, F_d comme fonctions de répartition marginales, la fonction définie par*

$$F^*(u_1, \dots, u_d) = \int \dots \int_{\mathbb{R}^d} I_{F_1}^*(F_1(x_1), u_1) \dots I_{F_d}^*(F_d(x_d), u_d) dF(x_1, \dots, x_d),$$

$(u_1, \dots, u_d) \in [0, 1]^d$, est une d -copule qui aux points de continuité de F_1, F_2, \dots, F_d vérifie,

$$F(x_1, \dots, x_d) = F^*(F_1(x_1), \dots, F_d(x_d)),$$

$(x_1, \dots, x_d) \in \mathbb{R}^d$.

Le théorème de [Niewiadomska-Bugaj et Kowalczyk, 2005] permet d'étendre l'utilisation des copules aux cas où les fonctions de répartition marginales des variables aléatoires ne sont pas nécessairement continues. Dans ces cas, on parle

plutôt de sous-copule et la transformation de [Szczesny, 1991] est un moyen de prolonger une sous-copule en une copule.

1.2. LES DIFFÉRENTS TYPES DE PROLONGEMENTS DE SOUS-COPULE

Dans le théorème de Sklar 1.0.1, la copule n'est définie de façon unique que sur un sous-ensemble de $[0, 1]^d$. En filigrane dans ce théorème, se dessine la notion de sous-copule.

Définition 1.2.1. *Une d -sous-copule D est une fonction qui satisfait les critères suivants :*

- *le domaine de définition de D est un sous-ensemble de $[0, 1]^d$, $R_1 \times \cdots \times R_d$, R_i étant un sous-ensemble de $[0, 1]$ contenant 0 et 1, $i = 1, \dots, d$;*
- *$D(u_1, \dots, u_d) = u_i$, si $u_1 u_2 \cdots u_d = u_i$, $u_i \in R_i$, $i = 1, \dots, d$;*
- *D est d -croissante c'est-à-dire pour $\mathbf{u} = (u_1, \dots, u_d)$, $\mathbf{v} = (v_1, \dots, v_d)$ tels que $u_1 \leq v_1, \dots, u_d \leq v_d$, $u_i, v_i \in R_i$, $i = 1, \dots, d$, le D -volume du rectangle $[\mathbf{u}, \mathbf{v}] = \mathbf{X}_{i=1}^d [u_i, v_i]$, $V_D([\mathbf{u}, \mathbf{v}])$ est positif ou nul, et est donné par*

$$V_D([\mathbf{u}, \mathbf{v}]) = \sum_{t \in \mathbf{X}_{i=1}^d \{u_i, v_i\}} (-1)^{n(t)} D(t),$$

où $n(t) = \#\{k : t_k = u_k, k = 1, \dots, d\}$, $t \in \mathbf{X}_{i=1}^d \{u_i, v_i\}$, \mathbf{X} étant mis pour le produit cartésien généralisé.

Une sous-copule peut être vue comme la restriction d'une copule sur un sous-ensemble de $[0, 1]^d$ de type $R_1 \times \cdots \times R_d$, R_i étant un sous-ensemble de $[0, 1]$ contenant 0 et 1. La relation entre une sous-copule et une copule est donnée par le théorème suivant.

Théorème 1.2.1. *Soit D une sous-copule définie sur $R_1 \times \cdots \times R_d$, R_i étant un sous-ensemble de $[0, 1]$ contenant 0 et 1. Il existe une copule C telle que $D(u_1, \dots, u_d) = C(u_1, \dots, u_d)$ pour tout $(u_1, \dots, u_d) \in R_1 \times \cdots \times R_d$; c'est-à-dire toute sous-copule peut être prolongée à une copule. Le prolongement n'est en général pas unique.*

Les sous-copules sont associées aux vecteurs aléatoires dont les fonctions de répartition marginales ne sont pas nécessairement continues. En effet, dans ce cas, les espaces images des fonctions de répartition marginales R_1, \dots, R_d , auxquels $\{0, 1\}$ est rajouté, ne sont pas toujours tous égaux à $[0, 1]$. Dans la littérature, il n'existe pas de famille de sous-copules, ce qui rend difficile l'utilisation des copules dans l'étude de la dépendance pour les vecteurs aléatoires dont les fonctions de répartition marginales sont discontinues. L'approche adoptée pour contourner le problème consiste à considérer les prolongements des sous-copules en copules [Denuit et Lambert, 2005], [Mesfioui et Quessy, 2010], [Genest et Nešlehová, 2007].

Dans cette thèse, nous ne considérerons que les copules non-paramétriques (définies comme des prolongements de sous-copules). En ce qui concerne l'utilisation des copules paramétriques pour l'analyse de telles variables, le lecteur peut se référer à [Nikoloulopoulos et Karlis, 2009], [McHale et Scarf, 2007], [Zimmer et Trivedi, 2006], [Cameron, Li, Trivedi et Zimmer, 2004], [Song, 2000], [Van Ophem, 1999]. Plusieurs techniques existent pour construire les prolongements de sous-copule. Pour cause de simplicité, nous donnons les prolongements en dimension 2.

1.2.1. Les prolongements analytiques

1.2.1.1. Le prolongement bilinéaire

Ce type de prolongement a été proposé par [Sklar, 1959], [Schweizer et Sklar, 1974], [Sklar, 1996].

Théorème 1.2.2. *Soit D une sous-copule, soit C la fonction définie par*

$$C(u_1, u_2) = \lim_{(x,y) \in R_1 \times R_2, (x,y) \rightarrow (u_1, u_2)} D(x, y)$$

pour tout $(u_1, u_2) \in \overline{R_1 \times R_2}$. Pour $(u_1, u_2) \notin \overline{R_1 \times R_2}$, définissons

$$\begin{aligned} a_1 &= \sup\{t \in \overline{R_1} : t \leq u_1\}, \\ b_1 &= \sup\{t \in \overline{R_2} : t \leq u_2\}, \\ a_2 &= \inf\{t \in \overline{R_1} : t \geq u_1\}, \\ b_2 &= \inf\{t \in \overline{R_2} : t \geq u_2\}, \end{aligned}$$

$$\begin{aligned} \lambda_1 &= \begin{cases} \frac{u_1 - a_1}{a_2 - a_1} & \text{si } a_1 < a_2 \\ 1, & \text{si } a_1 = a_2, \end{cases} \\ \lambda_2 &= \begin{cases} \frac{u_2 - b_1}{b_2 - b_1} & \text{si } b_1 < b_2 \\ 1, & \text{si } b_1 = b_2, \end{cases} \end{aligned}$$

$$C(u_1, u_2) = (1 - \lambda_1)(1 - \lambda_2)C(a_1, b_1) + (1 - \lambda_1)\lambda_2C(a_1, b_2) + \lambda_1(1 - \lambda_2)C(a_2, b_1) + \lambda_1\lambda_2C(a_2, b_2).$$

La fonction C est une copule qui prolonge D .

La preuve est tirée de [Nelsen, 1999, la preuve du lemme 2.3.5].

DÉMONSTRATION. Le domaine de la fonction C est $[0, 1]^2$ et C est égal à D sur $R_1 \times R_2$. Supposons sans perte de généralité que $\{0, 1\} \subset R_i$, $i = 1, 2$. Soient $u_1 = 0$, $u_2 \in [0, 1]$. Nous avons $\lambda_1 = 0$, $a_1 = 0$, $C(0, b_1) = 0$, $C(0, b_2) = 0$ et $C(0, u_2) = 0$. Par le même raisonnement, nous avons $C(u_1, 0) = 0$ pour $u_1 \in [0, 1]$. De façon similaire, pour $u_1 = 1$, $u_2 \in [0, 1]$, nous avons $\lambda_1 = 1$, $a_2 = 1$, $C(1, b_1) =$

$$b_1, C(1, b_2) = b_2,$$

$$C(1, u_2) = (1 - \lambda_2)b_1 + \lambda_2 b_2 = u_2.$$

De même, nous avons $C(u_1, 1) = u_1$ pour $u_1 \in [0, 1]$. Montrons maintenant que la fonction C est 2-croissante. Soient $(u_1, u_2), (u'_1, u'_2) \in [0, 1]^2$ tels que $u_1 \leq u'_1, u_2 \leq u'_2$. Posons

$$\begin{aligned} a_1 &= \sup\{t \in \overline{R_1} : t \leq u_1\}, \\ a'_1 &= \sup\{t \in \overline{R_1} : t \leq u'_1\}, \\ b_1 &= \sup\{t \in \overline{R_2} : t \leq u_2\}, \\ b'_1 &= \sup\{t \in \overline{R_2} : t \leq u'_2\}, \\ a_2 &= \inf\{t \in \overline{R_1} : t \geq u_1\}, \\ a'_2 &= \inf\{t \in \overline{R_1} : t \geq u'_1\}, \\ b_2 &= \inf\{t \in \overline{R_2} : t \geq u_2\}, \\ b'_2 &= \inf\{t \in \overline{R_2} : t \geq u'_2\}, \\ \lambda'_1 &= \begin{cases} \frac{u'_1 - a'_1}{a'_2 - a'_1} & \text{si } a'_1 < a'_2 \\ 1, & \text{si } a'_1 = a'_2, \end{cases} \\ \lambda'_2 &= \begin{cases} \frac{u'_2 - b'_1}{b'_2 - b'_1} & \text{si } b'_1 < b'_2 \\ 1, & \text{si } b'_1 = b'_2. \end{cases} \end{aligned}$$

Pour évaluer $V_C([u_1, u'_1] \times [u_2, u'_2])$, nous devons considérer plusieurs cas de figure. Si $a'_1 = a_1, b'_1 = b_1, a'_2 = a_2, b'_2 = b_2$, alors

$$\begin{aligned} V_C([u_1, u'_1] \times [u_2, u'_2]) &= (\lambda'_1 - \lambda_1)(\lambda'_2 - \lambda_2)V_C([a_1, a'_1] \times [b_2, b'_2]), \\ &= (\lambda'_1 - \lambda_1)(\lambda'_2 - \lambda_2)V_D([a_1, a'_1] \times [b_2, b'_2]) \geq 0. \end{aligned}$$

Pour les autres cas de figure, posons

$$\begin{aligned} e_1 &= \sup\{t \in \overline{R_1} : u_1 \leq t \leq u'_1\}, \\ f_1 &= \inf\{t \in \overline{R_1} : u_1 \leq t \leq u'_1\}, \\ e_2 &= \sup\{t \in \overline{R_2} : u_2 \leq t \leq u'_2\}, \\ f_2 &= \inf\{t \in \overline{R_2} : u_2 \leq t \leq u'_2\}, \end{aligned}$$

nous avons la décomposition suivante

$$[u_1, u'_1] \times [u_2, u'_2] = \{[u_1, e_1] \cup [e_1, f_1] \cup [f_1, u'_1]\} \times \{[u_2, e_2] \cup [e_2, f_2] \cup [f_2, u'_2]\}$$

Nous pouvons donc décomposer $[u_1, u'_1] \times [u_2, u'_2]$ en neuf rectangles R_1, \dots, R_9 (certains pouvant être vides); et pour ces rectangles, en nous inspirant du premier cas de figure, nous avons $V_C(R_i) \geq 0, i = 1, \dots, 9$. La fonction C est donc

2-croissante, à marges uniformes, s'annule dès que $u_1 = 0$ ou $u_2 = 0$ et coïncide avec D sur son domaine. C'est par conséquent un prolongement de D . \square

1.2.1.2. Les prolongements basés sur la méthode des rectangles

Les prolongements basés sur la méthode des rectangles sont une généralisation des prolongements bilinéaires. Des constructions de ce type ont été réalisées par [Mesiar et Szolgay, 2004], [Nelsen, Quesada-Molina, Rodríguez-Lallena et Úbeda-Flores, 2008], [Durante, Saminger-Platz et Sarkoci, 2009], [de Amo et coll., 2012]; la plus récente étant celle de [de Amo et coll., 2012]. La construction de [de Amo et coll., 2012] consiste à décomposer l'ensemble $[0, 1]^2$ en une famille de rectangles, puis à définir une mesure doublement stochastique à l'intérieur de chaque rectangle. Soit H une fonction de répartition conjointe en dimension 2 avec comme fonctions de répartition marginales F et G . Soit D la sous-copule associée à F définie sur $\overline{R_1 \cup \{0, 1\}} \times \overline{R_2 \cup \{0, 1\}}$ avec R_1, R_2 les ensembles images respectifs de F et G . Prenons l'ensemble $\overline{R_1 \cup \{0, 1\}}$ et considérons la famille S_1 des sous-intervalles fermés de $\overline{R_1 \cup \{0, 1\}}$ telle que leurs intersections deux à deux soient vides. Considérons la classe P_1 des éléments de S_1 qui sont des singletons, et posons $E_1 = S_1 \setminus P_1$. La fermeture des complémentaires dans $[0, 1]$ des éléments de S_1 est une famille notée O_1 . Soit T un ensemble indexant les éléments de $E_1 \cup O_1$, $E_1 \cup O_1 = \{[a_{1t}, a_{2t}], t \in T\}$. Réalisant la même décomposition pour $\overline{R_2 \cup \{0, 1\}}$, notons J l'ensemble indexant $E_2 \cup O_2$, $E_2 \cup O_2 = \{[b_{1j}, b_{2j}], j \in J\}$. Soient F_{tj} une fonction de répartition sur $[a_{1t}, a_{2t}]$ et G_{tj} une fonction de répartition sur $[b_{1j}, b_{2j}]$, $t \in T, j \in J$ satisfaisant les conditions de la partition de l'unité :

$$\begin{aligned} \sum_{j \in J} \frac{\beta_{tj}}{\sum_{j \in J} \beta_{tj}} F_{tj}(u_1) &= u_1, \quad u_1 \in [0, 1], \text{ pour } [a_{1t}, a_{2t}] \in O_1; \\ \sum_{t \in T} \frac{\beta_{tj}}{\sum_{t \in T} \beta_{tj}} G_{tj}(u_2) &= u_2, \quad u_2 \in [0, 1], \text{ pour } [b_{1j}, b_{2j}] \in O_2; \end{aligned}$$

$$\beta_{tj} = D(a_{1t}, b_{1j}) + D(a_{2t}, b_{2j}) - D(a_{1t}, b_{2j}) - D(a_{2t}, b_{1j}), \quad t \in T, j \in J.$$

Tout prolongement d'une sous-copule D par la méthode des rectangles est de la forme

$$\begin{aligned} C(u_1, u_2) &= D(a_{1t}, b_{1j}) + \beta_{tj} C_{tj}(F_{tj}(\frac{u_1 - a_{1t}}{a_{2t} - a_{1t}}), G_{tj}(\frac{u_2 - b_{1j}}{b_{2j} - b_{1j}})) + \sum_{t' \in S_t} \beta_{t'j} G_{t'j}(\frac{u_2 - b_{1j}}{b_{2j} - b_{1j}}) + \\ &\quad \sum_{j' \in Z_j} \beta_{tj'} F_{tj'}(\frac{u_1 - a_{1t}}{a_{2t} - a_{1t}}); \end{aligned}$$

$(u_1, u_2) \notin \overline{R_1 \cup \{0, 1\}} \times \overline{R_2 \cup \{0, 1\}}$, $(u_1, u_2) \in [a_{1t}, a_{2t}] \times [b_{1j}, b_{2j}]$, C_{tj} étant une copule, $S_t = \{t' : a_{1t'} < a_{1t}\}$ et $Z_j = \{j' : b_{1j'} < b_{1j}\}$, $t \in T, j \in J$. Pour tout

prolongement C , les fonctions F_{tj} et G_{tj} prennent la forme :

$$\begin{aligned} F_{tj}(u_1) &= \frac{1}{\beta_{tj}}(C((a_{2t} - a_{1t})u_1 + a_{1t}, b_{2j}) + C(a_{1t}, b_{1j}) - C(a_{1t}, b_{2j}) - \\ &\quad C((a_{2t} - a_{1t})u_1 + a_{1t}, b_{1j})); \\ G_{tj}(u_2) &= \frac{1}{\beta_{tj}}(C(a_{2t}, (b_{2j} - b_{1j})u_2 + b_{1j}) + C(a_{1t}, b_{1j}) - C(a_{2t}, b_{1j}) - \\ &\quad C(a_{1t}, (b_{2j} - b_{1j})u_2 + b_{1j})); \end{aligned}$$

$t \in T, j \in J$.

1.2.2. Les prolongements probabilistes

Certains des prolongements analytiques définis plus haut peuvent également être obtenus à partir de transformation sur des variables aléatoires. C'est ce qu'a réalisé Nešlehová [Nešlehová, 2007] à partir de la transformation de Szczesny [Szczesny, 1991]. Soit (X_1, X_2) un vecteur aléatoire de fonction de répartition conjointe H , la fonction de répartition marginale de la variable aléatoire X_1 étant F , la fonction de répartition marginale de la variable aléatoire X_2 étant G . Considérons les fonctions ψ_1, ψ_2

$$\begin{aligned} \mathbb{R} \times [0, 1] &\longrightarrow [0, 1] \\ (x_i, u_i) &\longmapsto \psi_i(x_i, u_i) = P(X_i < x_i) + u_i(P(X_i = x_i)), \end{aligned}$$

$i = 1, 2$. Soit U_i une variable aléatoire de loi uniforme sur $[0, 1]$ indépendante de $X_i, i \in \{1, 2\}$. Les variables aléatoires $\psi_1(X_1, U_1), \psi_2(X_2, U_2)$ sont de loi uniforme sur $[0, 1]$; la fonction de répartition conjointe de $(\psi_1(X_1, U_1), \psi_2(X_2, U_2))$ est une copule qui prolonge D . Si U_1 et U_2 sont deux variables aléatoires indépendantes, le prolongement obtenu est le prolongement bilinéaire. [Mesfioui et Quessy, 2010] ont généralisé cette transformation en dimension $d \geq 2$.

1.2.3. Le cas particulier de la copule empirique et de la copule empirique lissée

Ce fut [Deheuvels, 1979] qui introduisit pour la première fois la notion de copule empirique. Soient $(X_1, Y_1), \dots, (X_n, Y_n)$, n vecteurs aléatoires i.i.d, soient u_i le rang de X_i parmi X_1, \dots, X_n divisé par n , v_i le rang de Y_i parmi Y_1, \dots, Y_n divisé par $n, i = 1, \dots, n$. Toute copule \hat{C}_n dont la restriction sur $\{0, \frac{1}{n}, \frac{2}{n}, \dots, 1\}^2$ vérifie

$$\hat{C}_n\left(\frac{j_1}{n}, \frac{j_2}{n}\right) = \frac{1}{n} \sum_{i=1}^n I(u_i \leq \frac{j_1}{n}) I(v_i \leq \frac{j_2}{n}), \quad 0 \leq j_1, j_2 \leq n,$$

est une copule empirique. Un autre point de vue consiste à définir la sous-copule D ,

$$D\left(\frac{j_1}{n}, \frac{j_2}{n}\right) = \frac{1}{n} \sum_{i=1}^n I(u_i \leq \frac{j_1}{n}) I(v_i \leq \frac{j_2}{n}), \quad 0 \leq j_1, j_2 \leq n,$$

et à voir les copules empiriques comme des prolongements de la sous-copule D . Tout prolongement de D est un estimateur de la copule associée au couple (X_1, Y_1) . Le prolongement le plus utilisé dans ce cas est le prolongement bilinéaire.

Pour l'estimation, il est le plus souvent utilisé une pondération de la sous-copule D [voir [Sancetta et Satchell, 2004](#)]. En effet, considérons par exemple D' la sous-copule suivante :

$$D'(\frac{j_1}{n}, \frac{j_2}{n}) = \sum_{k_1=0}^n \sum_{k_2=0}^n D(\frac{k_1}{n}, \frac{k_2}{n}) p_{k_1,n}(\frac{j_1}{n}) p_{k_2,n}(\frac{j_2}{n}),$$

avec

$$p_{k,n}(u) = \binom{n}{k} u^k (1-u)^{n-k},$$

$u, v \in [0, 1]$, $0 \leq j_1, j_2 \leq n$, $k \in \{0, 1, \dots, n\}$. La sous-copule D' est une approximation de la sous-copule D :

$$|D'(\frac{j_1}{n}, \frac{j_2}{n}) - D(\frac{j_1}{n}, \frac{j_2}{n})| \leq \frac{8}{n},$$

$0 \leq j_1, j_2 \leq n$ (ceci est dû au fait que l'oscillation de D sur tout rectangle, dont les côtés sont de longueur $1/n$, est inférieure à $2/n$). Un prolongement de la sous-copule D' est la copule de Bernstein

$$\hat{C}(u, v) = \sum_{k_1=0}^n \sum_{k_2=0}^n D(\frac{k_1}{n}, \frac{k_2}{n}) p_{k_1,n}(u) p_{k_2,n}(v),$$

$$p_{k,n}(u) = \binom{n}{k} u^k (1-u)^{n-k},$$

$u, v \in [0, 1]$, $k \in \{0, 1, \dots, n\}$.

L'estimateur de la densité de copule est :

$$\hat{c}(u, v) = \sum_{k_1=0}^n \sum_{k_2=0}^n \hat{C}_n(\frac{k_1}{n}, \frac{k_2}{n}) p'_{k_1,n}(u) p'_{k_2,n}(v),$$

avec $u, v \in [0, 1]$, $p'_{k,n}$ est mis pour la dérivée de $p_{k,n}$, $k \in \{0, 1, \dots, n\}$. Ce sont des estimateurs bona fide de la copule et de la densité de copule. En dehors de ces estimateurs, on peut trouver les estimateurs basés sur la méthode du noyau [[Gijbels et Mielniczuk, 1990](#)], [[Fermanian, Radulovic et Wegkamp, 2002](#)], les estimateurs obtenus par décomposition sur une base de Riesz [[Genest, Masiello et Tribouley, 2009](#)]. Cependant pour ces estimateurs, l'estimateur de densité n'est pas une densité de copule.

1.3. BORNES DE FRÉCHET DES PROLONGEMENTS

1.3.1. Bornes de Fréchet des prolongements en dimension 2

Étant donnée une sous-copule bivariée D , les prolongements possibles ne sont pas uniques. Établir la borne inférieure LC et la borne supérieure des prolongements UC permet de connaître en chaque point $(u_1, u_2) \in [0, 1]^2$ l'intervalle $[LC(u_1, u_2), UC(u_1, u_2)]$ des valeurs admissibles pour les prolongements. Ce problème a été abordé pour la première fois par [Carley, 2002] dans le cas d'une sous-copule D avec un domaine de cardinalité finie. [Carley, 2002] a prouvé qu'en dimension 2, la borne supérieure et la borne inférieure étaient des copules et a donné leurs expressions. Soient $R_1 = \{a_{11}, a_{12}, \dots, a_{1n}\}$, $R_2 = \{a_{21}, a_{22}, \dots, a_{2m}\}$, $0 = a_{11} < a_{12} < \dots < a_{1n} = 1$, $0 = a_{21} < a_{22} < \dots < a_{2m} = 1$, la solution proposée par [Carley, 2002] est :

$$UC(u_1, u_2) = \sum_{j_1=2}^n \sum_{j_2=2}^m \max(0, \min(u_1 - a_{1j_1-1} - \sum_{l=1}^{j_2-1} p_{j_1 l}, u_2 - a_{2j_2-1} - \sum_{l=1}^{j_1-1} p_{l j_2}, p_{j_1 j_2}))$$

$$LC(u_1, u_2) = \sum_{j_1=2}^n \sum_{j_2=2}^m \max(0, \min(u_1 - a_{1j_1} + \sum_{l=1}^{j_2} p_{j_1 l}, p_{j_1 j_2}) + \min(u_2 - a_{2j_2} + \sum_{l=1}^{j_1} p_{l j_2}, p_{j_1 j_2}) - p_{j_1 j_2})$$

avec $p_{j_1 j_2} = D(a_{1j_1}, a_{2j_2}) - D(a_{1j_1-1}, a_{2j_2}) - D(a_{1j_1}, a_{2j_2-1}) + D(a_{1j_1-1}, a_{2j_2-1})$, $j_1 \in \{2, \dots, n\}$, $j_2 \in \{2, \dots, m\}$, $p_{11} = p_{1j_2} = p_{j_1 1} = 0$, $j_1 \in \{2, \dots, n\}$, $j_2 \in \{2, \dots, m\}$, $(u_1, u_2) \in [0, 1]^2$. Cependant, des difficultés apparaissent lors de la généralisation de cette formule à une dimension $d \geq 2$.

1.3.2. Bornes de Fréchet des prolongements en dimension $d > 2$

En dimension $d > 2$, il n'est pas toujours possible de construire une copule avec des marges multivariées fixées : ce problème est connu sous le nom de problème d'incompatibilité [Joe, 1997], [Nelsen, 1999], [Rodríguez-Lallena et Úbeda Flores, 2004], [Durante, Klement et Quesada-Molina, 2008]. Ce problème a été discuté par [Bass, 1955], [Genest, Quesada-Molina et Rodrigues-Lallena, 1995b], [Joe, 1997], des solutions partielles ont été proposées par [Bass, 1955], [Kellerer, 1964], [Dall'Aglio, 1972]. Les solutions apportées ne peuvent cependant s'appliquer aux prolongements de sous-copules ; en effet, des contraintes additionnelles s'imposent, des contraintes dont ne tiennent pas compte ces travaux. Mis à part, la tentative de [Carley, 2002] de trouver la borne supérieure des prolongements de sous-copules en dimension 3, à notre connaissance, il n'y a pas de travaux ayant réalisé une caractérisation de l'ensemble des prolongements d'une sous-copule en dimension $d > 2$ ou ayant établi la borne supérieure ou inférieure d'un tel ensemble.

1.4. APPORTS DE LA THÈSE ET CONCLUSION

La thèse fait des apports dans trois directions : l'estimation non paramétrique de la copule, la caractérisation de tous les prolongements d'une sous-copule donnée en dimension 2, et la recherche de la borne supérieure des prolongements d'une sous-copule en dimension 3.

En ce qui concerne l'estimation de la copule, il y a un manque de résultats relatifs à la convergence uniforme sur $[0, 1]^d$ et la convergence uniforme locale (sur un voisinage autour d'un point) de tels estimateurs. Le chapitre 2 donne le résultat local en s'appuyant sur le lissage spline de la copule empirique, tandis que le résultat sur $[0, 1]^d$ est donné pour l'estimateur copule de Bernstein empirique. L'apport principal du chapitre 2 est de fusionner les deux approches (locale et globale) dans un même cadre. C'est un premier pas vers la construction d'un estimateur multirésolution par insertion de nœuds à partir de la copule de Bernstein empirique. Un tel estimateur pourrait être utile par exemple pour l'étude de la dépendance locale (la dépendance entre un groupe de variables pour des niveaux fixés d'un autre groupe de variables), la dépendance spatiale (en utilisant comme paramètres de résolution des paramètres spatiaux), l'analyse multirésolution multivariée (construction d'ondelettes multivariées à partir de la copule empirique), l'étude de la dépendance pour les données de grandes dimensions (il est possible de diminuer la dimension du problème en remplaçant les données par les coefficients de l'analyse multirésolution). Le chapitre 3 s'attèle à caractériser l'ensemble des copules qui prolongent une sous-copule donnée en dimension 2 et donne la borne supérieure et la borne inférieure de cet ensemble. Ce travail a été par le passé réalisé par [de Amo et coll., 2012] mais la caractérisation proposée ne couvrait pas la dépendance dans les espaces très généraux. Considérons par exemple la sous-copule D définie sur $A \times [0, 1]$, $A \subset [0, 1]$, A étant nulle part dense de mesure de Lebesgue non nulle. La construction de [de Amo et coll., 2012] est inopérante dans ce cas. En effet, la construction de [de Amo et coll., 2012] utilise un nombre dénombrable de rectangles alors que le nombre d'éléments de A est non dénombrable. Le chapitre 3 illustre les limites de la construction de [de Amo et coll., 2012] dans le cas l'étude de la dépendance sur $[0, 1]^2$ en s'appuyant sur les ensembles de Cantor gras ; il donne la caractérisation la plus générale possible (théorème 3.2.1).

En dimension 3, il n'y a pas encore de travaux établissant la borne supérieure des prolongements d'une sous-copule. Le chapitre 4 essaie de combler ce vide. Il détermine la borne supérieure des prolongements d'une sous-copule de domaine

fini en dimension 3. Cette borne supérieure se trouve être une quasi-copule. Un exemple de la construction d'une telle borne est donné à la fin du chapitre.

Chapitre 2

CONVERGENCE DES ESTIMATEURS BONA FIDE DE COPULE ET DE DENSITÉ DE COPULE OBTENUS PAR LISSAGE SPLINE DE LA COPULE EMPIRIQUE

Résumé. *Nous proposons des estimateurs bona fide de la copule et de la densité de copule basés sur le lissage spline de la copule empirique. Ces estimateurs sont une généralisation de la copule de Bernstein empirique. Il y a un manque de résultats sur le comportement asymptotique des estimateurs de la densité de copule basés sur les pseudo-observations : le seul résultat connu est l'écart quadratique moyen (MISE) de l'estimateur par ondelettes [Genest et coll., 2009]. Nous donnons les résultats sur la convergence (presque sûre, quadratique) et la vitesse de convergence des estimateurs proposés.*

2.1. INTRODUCTION

Le théorème de Sklar 1.0.1 [Sklar, 1959], redécouvert par [Deheuvels, 1978], a ouvert la voie à l'utilisation des copules dans la modélisation multivariée. Étant donnée une fonction de répartition F d'un vecteur aléatoire (X_1, X_2, \dots, X_d) , avec comme fonctions de répartition marginales F_1, F_2, \dots, F_d , il existe une copule C telle que :

$$F(x_1, x_2, \dots, x_d) = C(F_1(x_1), F_2(x_2), \dots, F_d(x_d));$$

$(x_1, x_2, \dots, x_d) \in \mathbb{R}^d$. La densité de copule associée, c , lorsqu'elle existe, vérifie :

$$f(x_1, x_2, \dots, x_d) = c(F_1(x_1), F_2(x_2), \dots, F_d(x_d))f_1(x_1)f_2(x_2) \cdots f_d(x_d);$$

$(x_1, x_2, \dots, x_d) \in \mathbb{R}^d$, f est la densité conjointe du vecteur aléatoire (X_1, X_2, \dots, X_d) , f_1, f_2, \dots, f_d sont les densités marginales.

L'estimation de la copule ou de la densité de copule, lorsque les distributions marginales ne sont pas connues, n'est pas une tâche facile. Le plus souvent, l'inférence s'opère récursivement ou non en deux étapes ; une pour l'estimation des distributions marginales et une autre pour l'estimation de la copule ou de la densité de copule. À cet effet, on distingue les méthodes paramétriques (IFM, distributions marginales et copule estimées par maximum de vraisemblance [Joe, 2005]), semi-paramétrique (distributions marginales estimées par une méthode non paramétrique et copule estimée de façon paramétrique, [Genest, Ghoudi et Rivest, 1995a], [Shih et Louis, 1995], [Chen et Fan, 2006]). L'approche non paramétrique estime en une étape la copule en remplaçant les distributions marginales par les distributions empiriques correspondantes. Elle peut compléter utilement les approches paramétriques et semi paramétriques pour ce qui concerne la qualité de l'ajustement [Genest et coll., 2009]. Ce fut [Deheuvels, 1979] qui introduisit pour la première fois un estimateur non paramétrique de la copule qu'il appela copule empirique. Soient $\{(u_{11}, \dots, u_{d1}), \dots, (u_{1n}, \dots, u_{dn})\}$ les rangs associées à l'échantillon $\{(X_{11}, \dots, X_{d1}), \dots, (X_{1n}, \dots, X_{dn})\}$; toute copule \hat{C} dont la restriction définie sur $\{\frac{1}{n}, \frac{2}{n}, \dots, 1\}^d$ vérifie

$$\hat{C}(\frac{j_1}{n}, \dots, \frac{j_d}{n}) = \frac{1}{n} \sum_{i=1}^n I(u_{1i} \leq j_1) \cdots I(u_{di} \leq j_d), \quad 0 \leq j_1, \dots, j_d \leq n,$$

est une copule empirique. La copule empirique n'est pas unique. Néanmoins, il existe une approximation naturelle de la copule empirique,

$$\tilde{C}_n(u_1, \dots, u_d) = F_n(F_{1n}^-(u_1), \dots, F_{dn}^-(u_d)), \quad (u_1, \dots, u_d) \in [0, 1]^d,$$

selon la définition de [Deheuvels, 1979] avec F_n la fonction de répartition empirique,

$$F_n(x_1, \dots, x_d) = \frac{1}{n} \sum I(X_{1i} \leq x_1, \dots, X_{di} \leq x_d), \quad (x_1, \dots, x_d) \in \mathbb{R}^d,$$

$F_{1n}(\cdot) = F_n(\cdot, +\infty, \dots, +\infty), \dots, F_{dn}(\cdot) = F_n(+\infty, \dots, +\infty, \cdot)$ sont les fonctions de répartition marginales empiriques,

$$F_{jn}^-(u) = \inf\{t \in \mathbb{R}, F_{jn}(t) \geq u\}, \quad u \in [0, 1], \quad j = 1, \dots, d$$

sont les fonctions de quantile associées. Un autre estimateur défini par [Genest et coll., 1995a] est :

$$C_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^n I(F_{1n}(X_{1i}) \leq u_1, \dots, F_{dn}(X_{di}) \leq u_d), \quad (u_1, \dots, u_d) \in [0, 1]^d. \quad (2.1.1)$$

L'estimateur C_n présente l'inconvénient de ne pas être lisse. Différents lissages de la copule empirique et de l'histogramme associé à C_n ont été proposés parmi

lesquels le lissage par B-spline [Shen, Zhu et Song, 2008], par la méthode de noyau [Gijbels et Mielniczuk, 1990], [Fermanian et coll., 2002], par les ondelettes [Genest et coll., 2009], par polynômes de Bernstein [Bouezmarni, Rombouts et Taamouti, 2010], [Sancetta et Satchell, 2004], [Rödel, 1987]. L'estimateur par noyau de la densité de copule connaît des problèmes de biais multiplicatif et d'une vitesse de convergence du biais plus faible à la frontière du support [Hominal et Deheuvels, 1979], [Charpentier, Fermanian et Scaillet, 2007]. Des palliatifs ont été proposés : noyau image miroir [Hominal et Deheuvels, 1979], [Schuster, 1985], transformation bijective [Devroye et Györfi, 1985], noyau frontière [John, 1984], [Müller, 1991]. Un autre inconvénient de ces estimateurs est qu'ils ne sont pas des densités de copule car leurs distributions marginales ne sont pas uniformes. L'estimateur de la densité de copule par les polynômes de Bernstein présente l'avantage de ne pas avoir de biais multiplicatif et les estimateurs de la copule en découlant sont des copules. Une généralisation de l'estimation par polynômes de Bernstein est l'estimation par spline introduit par [Schoenberg, 1946]. Les splines ont l'avantage d'être des bases locales, et peuvent donner une meilleure vitesse de convergence [Cai et Meyer, 2011; Perron et Mengersen, 2001]. Nous généralisons les estimateurs par polynômes de Bernstein de la copule et de la densité de copule de [Sancetta et Satchell, 2004], déterminons la distance entre les nœuds donnant le meilleur compromis biais-variance, et donnons les résultats asymptotiques pour les estimateurs obtenus.

À la section 2 est présentée l'approximation spline d'une copule. De cette approximation sont tirés les estimateurs splines de la copule et de la densité de copule à la section 3. Les résultats asymptotiques y sont aussi présentés : ils s'appuient sur les résultats concernant le module d'oscillation d'un processus empirique [Stute, 1984], et sur les résultats asymptotiques de [Rüschendorf, 1974, 1976], [Einmahl et Ruymgaart, 1987; Deheuvels, 1997; Fermanian et coll., 2002]. Suivent respectivement aux sections 4 et 5, la comparaison avec les estimateurs existants et les démonstrations des résultats. La conclusion se trouve à la section 6.

2.2. APPROXIMATION DE COPULES PAR DES SPLINES

Nous construisons une approximation spline d'une copule basée sur des fonctions de répartition sur $[0, 1]^d$. Considérons des nœuds sur $(0, 1)$, $0 < x_{i1} \leq x_{i2} \leq \dots \leq x_{im_i} < 1$, $i = 1, \dots, d$. Nous supposons que

$$x_{ij} = \begin{cases} 0 & \text{pour } j = 0, -1, \dots, -k_i + 1 \\ 1 & \text{pour } j = m_i + 1, m_i + 2, \dots, m_i + k_i, \end{cases}$$

m_i entier naturel, $i = 1, 2, \dots, d$. Nous définissons les fonctions indicatrices $R_{ij}(x : k_i, m_i)$, $i = 1, 2, \dots, d$, $j = 1, \dots, m_i + k_i$ sur $(0, 1]$

$$R_{ij}(x : k_i, m_i) = \begin{cases} 1 & \text{si } \xi_{ij-k_i} < x \leq \xi_{ij-k_i+1}, \\ 0 & \text{sinon,} \end{cases}$$

où

$$\xi_{ij} = \frac{1}{k_i} \sum_{l=0}^{k_i-1} x_{ij+l},$$

$i=1, \dots, d$.

Considérons un vecteur aléatoire sur $(0, 1]^d$, (X_1, X_2, \dots, X_d) , avec comme fonction de répartition conjointe F et définissons

$$p_{k_1 \dots k_d j_1 \dots j_d} = E\left[\prod_{i=1}^d R_{ij_i}(X_i : k_i, m_i)\right], \quad 1 \leq j_i \leq m_i + k_i, \quad i = 1, \dots, d.$$

Soit C une copule, considérons la restriction de C sur $\prod_{i=1}^d (\xi_{ij_i-k_i}, \xi_{ij_i-k_i+1}]$:

$$C_{k_1 \dots k_d j_1 \dots j_d}(x_1, \dots, x_d) = \begin{cases} \frac{P(\xi_{1j_1-k_1} < X_1 \leq x_1, \dots, \xi_{dj_d-k_d} < X_d \leq x_d)}{P(\xi_{1j_1-k_1} < X_1 \leq x_1, \dots, \xi_{dj_d-k_d} < X_d \leq x_d)} & \text{si } p_{k_1 \dots k_d j_1 \dots j_d} > 0 \\ \frac{P(\xi_{1j_1-k_1} < X_1 \leq x_1, \dots, \xi_{dj_d-k_d} < X_d \leq x_d)}{P(\xi_{1j_1-k_1} < X_1 \leq \xi_{1j_1-k_1+1}) \times \dots \times P(\xi_{dj_d-k_d} < X_d \leq \xi_{dj_d-k_d+1})} & \text{sinon.} \end{cases}$$

La copule C prend la forme :

$$C(x_1, x_2, \dots, x_d) = \sum_{j_1=1}^{k_1+m_1} \sum_{j_2=1}^{k_2+m_2} \dots \sum_{j_d=1}^{k_d+m_d} p_{k_1 \dots k_d j_1 \dots j_d} C_{k_1 \dots k_d j_1 \dots j_d}(x_1, \dots, x_d), \quad (2.2.1)$$

$(x_1, \dots, x_d) \in (0, 1]^d$. Une approximation de C basée sur des fonctions de répartition sur $[0, 1]$, $\{G_{jk_i}\}_{j=1}^{m_i+k_i}$, $i = 1, \dots, d$, est donnée par :

$$G_{k_1 \dots k_d}(x_1, \dots, x_d) = \sum_{j_1=1}^{k_1+m_1} \dots \sum_{j_d=1}^{k_d+m_d} p_{k_1 \dots k_d j_1 \dots j_d} \prod_{i=1}^d G_{j_i k_i}(x_i).$$

Les splines polynomiales $\{G_{ji}, j = 1, \dots, m+i, i = 1, \dots, k_i\}$ sont construites selon les formules de récurrence :

$$G_{ji+1} = \begin{cases} H_{ji+1} + (1 - H_{ji+1})G_{ji} & \text{si } j = 1, \\ H_{ji+1}G_{j-1i} + (1 - H_{ji+1})G_{ji} & \text{si } j = 2, \dots, m+i, \\ H_{ji+1}G_{j-1i} & \text{si } j = m+i+1, \end{cases}$$

avec

$$H_{ji}(x) = \begin{cases} 0 & \text{si } x < x_{j-i}, \\ \frac{x-x_{j-i}}{x_j-x_{j-i}} & \text{si } x \in I_{ij}, \quad x_{j-i} \leq x < x_j, \\ I(x \geq x_j) & \text{si } x_{j-i} = x_j, \\ 1 & \text{si } x > x_j, \end{cases}$$

$i = 1, \dots, k_i$, et $G_{j0} = I(x \geq x_j)$, $j = 1, \dots, m$. Il est montré en annexe A que les fonctions $\{G_{jk_i}\}_{j=1}^{m_i+k_i}$, $i = 1, \dots, d$ sont des fonctions de répartition associées à des M-splines et qu'en particulier

$$N_{j_i-k_i-1, k_i+1} = G_{j_i-1, k_i} - G_{j_i, k_i},$$

est une B-spline, $j = 1, \dots, m_i + k_i$, $i = 1, \dots, d$. L'approximation $G_{k_1 \dots k_d}$ est une fonction de répartition [Schoenberg, 1967] car étant une combinaison convexe de fonctions de répartition. De plus, les marges sont uniformes grâce au lemme 2.2.1.

Lemme 2.2.1. *Les fonctions de répartition $\{G_{jk_i}\}_{j=1}^{m_i+k_i}$, $i = 1, \dots, d$ satisfont la condition*

$$\sum_{j=1}^{m_i+k_i} \frac{(x_{k_i j} - x_{k_i j-k_i})}{k_i} G_{jk_i}(x) = x, \quad x \in (0, 1], \quad i = 1, \dots, d$$

et $G_{k_1 \dots k_d}$ est une copule.

DÉMONSTRATION. D'après [Beutel, Gonska, Kacsó et Tachev, 2002, propriété P1], nous avons la propriété suivante pour les B-splines normalisés :

$$\sum_{j=1}^{m_i+k_i+1} \xi_{ij-k_i} N_{j-k_i-1, k_i+1}(x) = x, \quad 0 \leq x \leq 1,$$

avec

$$\xi_{ij} = \frac{1}{k_i} \sum_{l=0}^{k_i-1} x_{ij+l},$$

$j=1, \dots, k+m$, $i=1, \dots, d$. Notons que nous avons

$$\xi_{ij-k_i+1} - \xi_{ij-k_i} = \frac{(x_{k_i j} - x_{k_i j-k_i})}{k}, \quad j = 1, \dots, k+m, \quad i = 1, \dots, d.$$

Pour $x \in (0, 1]$, $i = 1, \dots, d$, nous avons

$$\begin{aligned} \sum_{j=1}^{m_i+k_i} \frac{(x_{k_i j} - x_{k_i j-k_i})}{k_i} G_{jk_i}(x) &= \sum_{j=1}^{m_i+k_i} (\xi_{ij-k_i+1} - \xi_{ij-k_i}) G_{jk_i}(x), \\ &= \sum_{j=1}^{m_i+k_i} \xi_{ij-k_i+1} G_{jk_i}(x) - \sum_{j=1}^{m_i+k_i} \xi_{ij-k_i} G_{jk_i}(x), \\ &= \sum_{j'=2}^{m_i+k_i+1} \xi_{ij'-k_i} G_{j'-1, k_i}(x) - \sum_{j=1}^{m_i+k_i} \xi_{ij-k_i} G_{jk_i}(x), \\ &\quad \text{avec } j' = j+1, j = 1, \dots, m_i+k_i, \\ &= \sum_{j'=2}^{m_i+k_i+1} \xi_{ij'-k_i} G_{j'-1, k_i}(x) - \sum_{j=2}^{m_i+k_i} \xi_{ij-k_i} G_{jk_i}(x) \\ &\quad - \xi_{i1-k_i} G_{1, k_i}(x), \\ &= \sum_{j'=2}^{m_i+k_i} \xi_{ij'-k_i} (G_{j'-1, k_i}(x) - G_{j', k_i}(x)) + \xi_{im_i+1} G_{m_i+k_i, k_i}(x) \\ &\quad - \xi_{i1-k_i} G_{1, k_i}(x). \end{aligned}$$

Comme $G_{j'-1k_i} - G_{j'k_i} = N_{j'-k_i-1, k_i+1}$, $j' = 2, \dots, m_i + k_i + 1$ (lemme A.1.1), $\xi_{i1-k_i} = 0$ (de part la définition de ξ_{i1-k_i} les éléments de la sommation la définissant sont tous égaux à 0, $x_{1-k_i} = \dots = x_0 = 0$), $i = 1, \dots, d$, nous pouvons écrire

$$\begin{aligned} \sum_{j=1}^{m_i+k_i} \frac{(x_{k_i j} - x_{k_i j - k_i})}{k_i} G_{jk_i}(x) &= \sum_{j'=2}^{m_i+k_i} \xi_{ij'-k_i} N_{j'-k_i-1, k_i+1}(x) \\ &+ \xi_{im_i+1} N_{m_i, k_i+1}(x) + \xi_{i1-k_i} N_{-k_i, k_i+1}(x), \\ &= \sum_{j'=1}^{m_i+k_i+1} \xi_{ij'-k_i} N_{j'-k_i-1, k_i+1}(x), \\ &= x, \end{aligned}$$

$x \in (0, 1]$. □

À l'issue de cette démonstration, on a pu s'apercevoir que G_k pouvait aussi s'écrire à l'aide de B -splines,

$$G_k(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} C(\xi_{j_1}, \dots, \xi_{j_d}) \prod_{i=1}^d N_{j_i-k-1, k+1}(u_i),$$

$(u_1, \dots, u_d) \in [0, 1]^d$, en utilisant l'équation A.1.3

$$G_{j, i} = \sum_{l=j}^{m+i} N_{l-i, i+1},$$

$j = 1, \dots, m + i$, $i = 1, \dots, k$.

Le lemme qui suit donne une idée de la qualité de l'approximation obtenue.

Lemme 2.2.2. *Nous avons*

$$\|G_{k_1 \dots k_d} - C\|_\infty \leq \sum_{i=1}^d \omega_1(C, h_1, \dots, h_d) (1 + \min\{\frac{1}{h_i \sqrt{(2k_i - 2)}}, \frac{\sqrt{\frac{k_i}{12}} |\Delta_i|}{h_i}\}), \quad (2.2.2)$$

$$\|G_{k_1 \dots k_d} - C\|_\infty \leq \sum_{i=1}^d \omega_2(C, h_1, \dots, h_d) (1 + \frac{1}{2h_i^2} \min\{\frac{1}{2k_i}, \frac{(k_i + 1) |\Delta_i|^2}{12}\})),$$

avec

$$\omega_r(C, h_1, \dots, h_d) = \sup_{|h_1^*| < h_1, \dots, |h_d^*| < h_d, x \in [0, 1]^d} \left\| \sum_{l=0}^r (-1)^{r-l} \binom{r}{l} C(x + (r-2l) \frac{h_l^*}{2}) \right\|$$

$|\Delta_i|$ est la distance maximale entre deux nœuds distincts consécutifs sur l'axe i , $i = 1, 2, \dots, d$, $h_1, \dots, h_d \in [0, 1]$.

DÉMONSTRATION. Ces inégalités ont été tirées de [Marsden, 1972; Beutel et coll., 2002]. □

L'approximation est d'ordre $O(|\Delta|/k)$ pour $k = k_1 = \dots = k_d$, $|\Delta_1| = \dots = |\Delta_d| = |\Delta|$, $h_1 = \dots = h_d = |\Delta|/k$, lorsque C admet des dérivées partielles d'ordre 1 continues. L'approximation est d'ordre $O(|\Delta|^2/k)$ pour $h_1 = \dots = h_d = |\Delta|/\sqrt{k}$ si C admet des dérivées partielles d'ordre 2 continues (voir [Beutel et coll., 2002]).

2.3. ESTIMATION DE LA COPULE ET DE LA DENSITÉ DE COPULE : PRINCIPAUX RÉSULTATS

Pour la suite, nous utilisons des nœuds uniformes tels que $n \frac{(x_{kj_i} - x_{kj_i-k})}{k}$ soit entier, $j_i = 1, \dots, k+m$, $i = 1, \dots, d$, n désignera le nombre d'observations. Cette hypothèse permet de simplifier les notations et les calculs (mais elle n'est pas indispensable) et elle nous permettra d'écrire

$$\frac{\#\{l : n\xi_{j_i-k}^l < R_i^l \leq n\xi_{j_i-k+1}^l; l = 1, \dots, n\}}{n} = \xi_{j_i-k+1} - \xi_{j_i-k} = \frac{(x_{kj_i} - x_{kj_i-k})}{k},$$

$j_i = 1, \dots, k+m$, $\{(R_1^1, \dots, R_1^n), \dots, (R_d^1, \dots, R_d^n)\}$ les vecteurs de rang associées à l'échantillon $\{(X_{11}, \dots, X_{1d}), \dots, (X_{n1}, \dots, X_{nd})\}$.

Comme le terme $\frac{(x_{kj_i} - x_{kj_i-k})}{k}$, $j_i = 1, \dots, k+m$, $i = 1, \dots, d$, jouera un rôle important dans les preuves, nous introduirons l'expression

$$\frac{|\pi|}{k} = \max_{1 \leq j_i \leq k+m} \frac{(x_{j_i} - x_{j_i-k})}{k}.$$

Cette expression $|\pi|/k$ sera l'analogue de la fenêtre dans une estimation par noyau. Dans le cas de l'estimation par polynômes de Bernstein, l'expression $|\pi|/k$ revient à $1/k$, et dans l'estimation par splines, k est fixé et $|\pi|/k$ est la distance maximale entre k nœuds consécutifs divisée par une constante k . Nous aurions pu prendre $|\Delta|/k$ mais $|\pi|/k$ a été choisi parce qu'il rend les calculs aisés et correspond grossièrement à la distance entre deux points de quasi-interpolation.

2.3.1. Estimation de la copule

Soit C une copule et C_n la copule empirique associée (2.1.1), l'estimateur que nous proposons est une approximation spline 2.2.1 avec $k = k_1 = \dots = k_d$, $m_1 = \dots = m_d$,

$$G_k(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} C_n(\xi_{j_1}, \dots, \xi_{j_d}) \prod_{i=1}^d N_{j_i-k-1, k+1}(u_i),$$

où C_n est la copule empirique (2.1.1), $N_{j_i-k-1, k+1}$ est une B-spline vérifiant

$$N_{j_i-k-1, k+1} = G_{j_i-1, k} - G_{j_i, k},$$

$u_i \in [0, 1]$, $i = 1, \dots, d$; G_k prend aussi la forme

$$G_k(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} p_{k, j_1 \dots j_d} \prod_{i=1}^d G_{j_i k}(u_i),$$

$$p_{k, j_1 \dots j_d} = \frac{\#\{l : n\xi_{j_1-k} < R_1^l \leq n\xi_{j_1-k+1}, \dots, n\xi_{j_d-k} < R_d^l \leq n\xi_{j_d-k+1}; l = 1, \dots, n\}}{n},$$

$$j_i = 1, \dots, m + k(n), u_i \in [0, 1], i = 1, \dots, d, l = 1, \dots, n.$$

Lemma 2.3.1. *L'estimateur G_k est une copule. Pour $k = 1$, G_k est l'estimateur empirique B-spline de [Shen et coll., 2008].*

DÉMONSTRATION. Les fonctions de répartition marginales de G_k sont uniformes d'après le lemme 2.2.1

$$G_k(1, \dots, 1, u_i, 1, \dots, 1) = \sum_{j_i=1}^{k+m} \frac{(x_{kj_i} - x_{kj_i-k})}{k} G_{j_i k}(u_i) = u_i, u_i \in [0, 1], i = 1, \dots, d.$$

Ceci est dû aux propriétés de partition de l'unité des B-splines. Pour $k = 1$, nous sommes en présence de B-splines linéaires avec nœuds uniformes [Shen et coll., 2008]. \square

La copule G_k peut être décomposée en deux termes : la copule indépendante plus un reste. Cette propriété est aussi vérifiée par la copule de Bernstein et peut être utilisée pour tester l'indépendance.

Lemma 2.3.2. *Nous avons la représentation suivante de la copule G_k*

$$G_k(u_1, \dots, u_d) = \prod_{i=1}^d u_i + \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} (p_{k, j_1 \dots j_d} - \prod_{i=1}^d (\xi_{ij_i-k+1} - \xi_{ij_i-k})) \prod_{i=1}^d G_{j_i k}(u_i),$$

$i = 1, \dots, d$. En particulier, pour $m = 0$, G_k est la copule de Bernstein de [San-cetta et Satchell, 2004] avec les pseudo-observations à la place des observations.

DÉMONSTRATION. Ceci est dû au fait que

$$\sum_{j_i=1}^{k+m} \frac{(x_{kj_i} - x_{kj_i-k})}{k} G_{j_i k}(u_i) = u_i, u_i \in [0, 1], i = 1, \dots, d.$$

et

$$\xi_{ij_i-k+1} - \xi_{ij_i-k} = \frac{(x_{kj_i} - x_{kj_i-k})}{k}, j_i = 1, \dots, k + m, i = 1, \dots, d.$$

\square

Les deux théorèmes qui suivent donnent le comportement asymptotique presque sûr et la convergence en loi de l'estimateur de la copule. Le théorème 2.3.1 donne une condition suffisante pour une convergence uniforme de l'ordre de $\sqrt{\frac{\log \log n}{n}}$. Cette condition s'appuie sur la majoration du lemme 2.2.2.

Théorème 2.3.1. *Supposons que la copule C a des dérivées partielles continues d'ordre 1. Soit $\{\varphi(n), n \geq 1\}$ une suite de nombre réels positifs telle que*

$$\lim_{n \rightarrow \infty} \varphi(n) \sqrt{\frac{\log \log n}{n}} = 0$$

alors

$$\lim_{n \rightarrow \infty} \varphi(n) \|G_k - C\|_\infty = 0$$

presque sûrement si

$$\min\left\{\frac{1}{\sqrt{(2k-2)}}, \sqrt{\frac{k}{12}}|\Delta|\right\} = O\left(\sqrt{\frac{\log \log n}{n}}\right),$$

$|\Delta|$ la distance maximale séparant deux nœuds consécutifs distincts.

La condition

$$\min\left\{\frac{1}{\sqrt{(2k-2)}}, \sqrt{\frac{k}{12}}|\Delta|\right\} = O\left(\sqrt{\frac{\log \log n}{n}}\right)$$

donne pour $|\Delta|$ petit, par exemple $|\Delta| = 1/\sqrt{n}$, $k = O(\log \log n)$; cette condition devient $k = O(1)$ pour $|\Delta| = \sqrt{\log \log n/n}$. Pour des valeurs élevées de $|\Delta|$, par exemple $|\Delta| = 1$, nous avons $k = O(n/\log \log n)$.

DÉMONSTRATION. Preuve du théorème 2.3.1

Définissons Q_π comme l'opérateur de quasi-interpolation par les splines de [Schoenberg, 1946] : Q_π est aussi un opérateur de projection. En utilisant l'inégalité triangulaire, nous obtenons la décomposition suivante :

$$\begin{aligned} \|G_k - C\|_\infty &\leq \|G_k - Q_\pi C\|_\infty + \|Q_\pi C - C\|_\infty, \\ &\leq \|Q_\pi(C_n - C)\|_\infty + \|Q_\pi C - C\|_\infty, \\ &\leq \|C_n - C\|_\infty + \|Q_\pi C - C\|_\infty, \end{aligned}$$

Pour $\|C_n - C\|_\infty$, la preuve de la convergence presque sûre d'ordre $O(\sqrt{\frac{\log \log n}{n}})$ a été faite par [Deheuvels, 1979] :

$$\lim_{n \rightarrow \infty} \varphi(n) \|C_n - C\|_\infty = 0.$$

Le résultat utilise le lemme 2.2.2 : si la condition

$$\min\left\{\frac{1}{\sqrt{(2k-2)}}, \sqrt{\frac{k}{12}}|\Delta|\right\} = O\left(\sqrt{\frac{\log \log n}{n}}\right)$$

est respectée, alors

$$\lim_{n \rightarrow \infty} \varphi(n) \|Q_\pi C - C\|_\infty = 0$$

et le résultat est acquis. □

Le processus $\{\sqrt{n}(C_n - C)(u_1, \dots, u_d) : (u_1, \dots, u_d) \in [0, 1]^d\}$ converge en loi vers un pont brownien arrêté. Ce résultat est aussi valable pour le processus $\{\hat{G}_k(u_1, \dots, u_d) = \sqrt{n}(G_k - C)(u_1, \dots, u_d) : (u_1, \dots, u_d) \in [0, 1]^d\}$.

Théorème 2.3.2. *Supposons que la copule C admet des dérivées partielles d'ordre 2, le processus*

$$\{\hat{G}_k(u_1, \dots, u_d) = \sqrt{n}(G_k - C)(u_1, \dots, u_d) : (u_1, \dots, u_d) \in [0, 1]^d\}$$

converge dans $l^\infty[0, 1]$ vers un processus gaussien $(G(u_1, \dots, u_d))_{(u_1, \dots, u_d) \in [0, 1]^d}$:

$$G(u_1, \dots, u_d) = B(u_1, \dots, u_d) - \sum_{l=1}^d \frac{\partial C(u_1, \dots, u_l, \dots, u_d)}{\partial u_l} B(1, \dots, 1, u_l, 1, \dots, 1)$$

si

$$\sqrt{n} \min\left\{\frac{1}{2k}, \frac{k+1}{12} |\Delta|^2\right\} = o(1).$$

avec $|\Delta|$ la distance maximale séparant deux nœuds consécutifs distincts,

$(B(u_1, \dots, u_d))_{(u_1, \dots, u_d) \in [0, 1]^d}$ un pont brownien.

La démonstration s'appuie sur la représentation presque sûre du processus empirique $\{\sqrt{n}(C_n - C)(u_1, \dots, u_d) : (u_1, \dots, u_d) \in [0, 1]^d\}$, $n > 0$.

DÉMONSTRATION. Preuve du théorème 2.3.2

Le processus $\sqrt{n}(C_n(u, v) - C(u, v))$ converge en loi vers le processus gaussien $(G_{(u,v)})_{(u,v) \in [0, 1]^2}$ d'après [Fermanian et coll., 2002]. D'après la représentation forte du processus empirique de copule de [Zari, 2010], nous pouvons trouver une version continue de G telle que $\sqrt{n}(C_n - C) - G$ tende presque sûrement vers 0 lorsque n tend vers l'infini. Soit Q_π l'opérateur consistant à lisser par les splines. Nous pouvons donner une représentation forte du processus lissé par la copule en faisant la décomposition suivante :

$$\sqrt{n}(G_k - C) - G = \sqrt{n}(G_k - Q_\pi C) - Q_\pi G + (Q_\pi G - G) + \sqrt{n}(Q_\pi C - C).$$

En appliquant la norme uniforme et l'inégalité triangulaire, la condition suffisante pour que $\|\sqrt{n}(G_k - C) - G\|_\infty$ tende vers 0 est $\sqrt{n}|\Delta|(1 + \min\{\frac{1}{|\Delta|\sqrt{(2k-2)}}, \sqrt{\frac{k}{12}}\}) = o(1)$ car les termes

$$\|\sqrt{n}(G_k - Q_\pi C) - Q_\pi G\|_\infty = \|Q_\pi(\sqrt{n}(C_n - C) - G)\|_\infty \leq \|\sqrt{n}(C_n - C) - G\|_\infty,$$

(d'après l'équation 2.2.2)

$$\|Q_\pi G - G\|_\infty \rightarrow 0, \quad \Delta \rightarrow 0,$$

$$\sqrt{n}|Q_\pi C - C| \leq \sqrt{n}|\Delta|(1 + \min\{\frac{1}{|\Delta|\sqrt{(2k-2)}}, \sqrt{\frac{k}{12}}\}).$$

(d'après l'équation 2.2.2, en prenant $h_1 = \dots = h_d = \Delta$). Nous rappelons que l'équation 2.2.2 est

$$\|G_{k_1 \dots k_d} - C\|_\infty \leq \sum_{i=1}^d \omega_1(C, h_1, \dots, h_d) (1 + \min\{\frac{1}{h_i \sqrt{(2k_i - 2)}}, \frac{\sqrt{\frac{k_i}{12}} |\Delta_i|}{h_i}\}).$$

Rappelons que $|\Delta|$ est la distance maximale entre deux nœuds consécutifs distincts. \square

2.3.2. Estimation de la densité de copule

De l'estimateur de la copule, nous tirons un estimateur de la densité de copule. L'estimateur spline de la densité de copule, g_k , est donné par la formule :

$$g_k(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} p_{k, j_1 \dots j_d} \prod_{i=1}^d M_{j_i-k, k}(u_i),$$

avec

$$M_{j_i-k, k}(u_i) = G'_{j_i k}(u_i), \quad j_i = 1, \dots, k+m, i = 1, \dots, d$$

$u_i \in (0, 1)$, $i = 1, \dots, d$. Avant d'énoncer les résultats de cette section, nous avons besoin de donner des définitions et résultats préliminaires.

Soient $U_{i1} = F_i(X_{i1}), U_{i2} = F_i(X_{i2}), \dots, U_{in} = F_i(X_{in})$ des variables i.i.d de loi uniforme sur $[0, 1]$, X_{i1}, \dots, X_{in} des variables aléatoires réelles de fonction de répartition F_i , $i = 1, \dots, d$. Le vecteur (U_{1l}, \dots, U_{dl}) a pour fonction de répartition une copule C de fonction de densité c , $l = 1, \dots, n$. Définissons le processus empirique

$$\alpha_n(\prod_{i=1}^d [0, t_i]) = \sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n [\prod_{i=1}^d I_{\{U_{ik} \leq t_i\}} - C(t)] \right),$$

$t = (t_1, \dots, t_d) \in [0, 1]^d$. Un résultat de [Stute, 1984] donne la vitesse de convergence du module d'oscillation de α_n

$$\omega_n(h, t) = \sup\{|\alpha_n(R)|, R = (t_1, t_1 + h] \times \dots \times (t_d, t_d + h]\},$$

$$t = (t_1, \dots, t_d) \in [0, 1]^d,$$

$$\lim_{n \rightarrow \infty} \frac{\omega_n(h_n, t)}{\sqrt{2h_n^d \log(h_n^d)}} = \sqrt{\sup_{u \in [0, 1]^d} c(u)};$$

$c(u) \geq m > 0$; la valeur $\sqrt{\sup_{u \in [0, 1]^d} c(u)}$ peut être remplacée par $\sqrt{\sup_{u \in R_0} c(u)}$, si tous les rectangles R sont inclus dans R_0 . La suite h_n , $n > 0$ doit vérifier les conditions

$$\begin{aligned} \lim_{n \rightarrow \infty} n(h_n)^d &= \infty, \quad \lim_{n \rightarrow \infty} h_n = 0, \\ \lim_{n \rightarrow \infty} \log(h_n^{-d}) &= o(nh_n^d), \end{aligned}$$

$$\lim_{n \rightarrow \infty} \log(h_n^{-d})(\log \log n)^{-1} = \infty.$$

Lemme 2.3.1. *Posons*

$$\alpha'_n(\prod_{i=1}^d [0, t_i]) = \sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n \prod_{i=1}^d I_{\{F_{in}(X_{ik}) \leq t_i\}} - C(t) \right),$$

F_{in} étant la fonction de répartition empirique associée à X_{i1}, \dots, X_{in} , $i = 1, \dots, d$,
 $t \in [0, 1]^d$.

Nous avons

$$|(\alpha'_n - \alpha_n)(\prod_{i=1}^d [0, t_i])| \leq \omega_n(\epsilon, t) + \omega_n(-\epsilon, t) + \sqrt{n}[C(t + \epsilon(1, \dots, 1)) - C(t - \epsilon(1, \dots, 1))],$$

$$\text{avec } \epsilon = \sqrt{2 \frac{\log \log n}{n}}.$$

DÉMONSTRATION. En considérant la loi du logarithme itéré, nous avons

$$|F_{in}(X_{ik}) - F_i(X_{ik})| \leq \sqrt{2 \frac{\log \log n}{n}},$$

$i = 1, \dots, d$ presque sûrement pour n grand. Ceci implique que

$$\begin{aligned} \alpha'_n(\prod_{i=1}^d [0, t_i]) &\leq \sqrt{n} \left(\frac{1}{n} \sum_{k=1}^n \prod_{i=1}^d I_{\{F_i(X_{ik}) \leq t_i + \epsilon\}} - C(t) \right), \\ &\leq \alpha_n(\prod_{i=1}^d [0, t_i + \epsilon]) + \sqrt{n}(C(t + \epsilon(1, \dots, 1)) - C(t)), \\ &\leq \alpha_n(\prod_{i=1}^d [0, t_i + \epsilon]) + \sqrt{n}(C(t + \epsilon(1, \dots, 1)) - C(t - \epsilon(1, \dots, 1))). \end{aligned}$$

On peut donc en déduire que

$$\begin{aligned} (\alpha'_n - \alpha_n)(\prod_{i=1}^d [0, t_i]) &\leq \omega_n(\epsilon, t) + \sqrt{n}(C(t + \epsilon(1, \dots, 1)) - C(t - \epsilon(1, \dots, 1))), \\ &\leq \omega_n(\epsilon, t) + \omega_n(-\epsilon, t) + \sqrt{n}(C(t + \epsilon(1, \dots, 1)) - C(t - \epsilon(1, \dots, 1))). \end{aligned}$$

Nous pouvons écrire le même type d'inégalité avec $\alpha_n - \alpha'_n$ en remplaçant $t + \epsilon(1, \dots, 1)$ par $t - \epsilon(1, \dots, 1)$. Finalement, nous avons

$$|(\alpha'_n - \alpha_n)(\prod_{i=1}^d [0, t_i])| \leq \omega_n(\epsilon, t) + \omega_n(-\epsilon, t) + \sqrt{n}(C(t + \epsilon(1, \dots, 1)) - C(t - \epsilon(1, \dots, 1))).$$

□

Posons

$$f_n(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} V_C(A_{j_1 \cdots j_d}) \prod_{l=1}^d M_{j_l-k, k}(u_l),$$

$$\begin{aligned} T_n(u_1, \dots, u_d) &= \frac{1}{n} \sum_{i=1}^n \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} I_{A_{j_1 \cdots j_d}}(F_1(X_{1i}), \dots, F_d(X_{di})) \prod_{l=1}^d M_{j_l-k, k}(u_l), \\ &= \frac{1}{\sqrt{n}} [\sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} \alpha_n(A_{j_1 \cdots j_d}) \prod_{l=1}^d M_{j_l-k, k}(u_l)] + f_n(u_1, \dots, u_d), \end{aligned}$$

l'estimateur g_k s'écrit

$$g_k(u_1, \dots, u_d) = \frac{1}{\sqrt{n}} \left[\sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} \alpha'_n(A_{j_1 \dots j_d}) \prod_{i=1}^d M_{j_i-k, k}(u_i) \right] + f_n(u_1, \dots, u_d),$$

$(u_1, \dots, u_d) \in [0, 1]^d$, V_C est la fonction volume associée à la copule C , $A_{j_1 \dots j_d} = \prod_{l=1}^d [\xi_{j_l-k}, \xi_{j_l-k+1})$, $j_i = 1, \dots, k+m$, $i = 1, \dots, d$. Les résultats asymptotiques présentés aux théorèmes 2.3.3 et 2.3.4 sont essentiellement ceux de T_n ; ces résultats sont identiques à ceux de g_k si $g_k - T_n$ décroît assez vite par rapport à $T_n - f_n$ ou est du même ordre; une condition suffisante pour que ce soit le cas est que h_n décroît moins vite que ϵ (en effet si la densité c est bornée au voisinage de t , on peut borner $C(t + \epsilon(1, \dots, 1)) - C(t - \epsilon(1, \dots, 1))$ par une constante multipliée par ϵ). Les conditions posées (K1) – (K4) ne sont qu'une adaptation des conditions de [Stute, 1984] avec $h_n = |\pi|/k$ auxquelles nous avons rajouté la condition (K2) pour que $g_k - T_n$ puisse décroître assez vite par rapport à $T_n - f_n$ pour n grand. La condition (K5) permet de s'assurer que la variance de g_k et la variance de T_n , $n > 0$, sont du même ordre.

Théorème 2.3.3. *Supposons que k ou $|\pi|$ soit fixé, et que la densité de la copule dans un voisinage de (u_1, \dots, u_d) soit finie et continue. Supposons que les conditions suivantes sont satisfaites :*

$$(K1) \quad \lim_{n \rightarrow \infty} n \left(\frac{|\pi|}{k} \right)^d = \infty, \quad \lim_{n \rightarrow \infty} \frac{|\pi|}{k} = 0,$$

$$(K2) \quad \lim_{n \rightarrow \infty} \left(\frac{|\pi|}{k} \right)^d (\epsilon)^{-1} \in (a, \infty], a > 0,$$

$$(K3) \quad \lim_{n \rightarrow \infty} \log(h_n^{-d}) = o(nh_n^d),$$

$$(K4) \quad \lim_{n \rightarrow \infty} \log(h_n^{-d})(\log \log n)^{-1} = \infty$$

$$(K5) \quad \lim_{n \rightarrow \infty} \left(\frac{|\pi|}{k} \right)^{d-d_1/2} \log n \in [0, c], c > 0$$

Le biais et la variance de l'estimateur g_k vérifient :

$$|E[(g_k - c)(u_1, \dots, u_d)]| = O(\omega_1(c, \frac{|\pi|}{k}, \dots, \frac{|\pi|}{k}));$$

$$\text{Var}(g_k - c)(u_1, \dots, u_d) = O\left(\frac{c(u_1, \dots, u_d)}{n \left(\frac{|\pi|}{k} \right)^{d-d_1/2}}\right), \text{ si } |\pi| \text{ est fixé};$$

$$\text{Var}(g_k - c)(u_1, \dots, u_d) = O\left(\frac{c(u_1, \dots, u_d)}{n \left(\frac{|\pi|}{k} \right)^d}\right), \text{ si } k \text{ est fixé};$$

d_1 est le nombre de u_1, \dots, u_d différents de 0 et 1. Le choix optimal pour $\frac{|\pi|}{k}$ par rapport à l'écart quadratique moyen (EQM) en (u_1, \dots, u_d) est

$$\frac{|\pi|}{k} = O(n^{-\frac{1}{2+d-d_1/2}}), \text{ si } |\pi| \text{ est fixé};$$

La vitesse de convergence en probabilité pour ce choix est $n^{-\frac{1}{2+d-d_1/2}}$ pour $|\pi|$ fixé.

DÉMONSTRATION. L'ensemble des suites satisfaisant aux conditions (K1) – (K5) n'est pas vide. Par exemple, le choix

$$\frac{|\pi|}{k}^d = \frac{\log n}{\sqrt{n}}$$

satisfait les conditions (K1) – (K5). La démonstration consiste à montrer principalement que

$$|E[(T_n - c)(u_1, \dots, u_d)]| = O(\omega_1(c, \frac{|\pi|}{k}, \dots, \frac{|\pi|}{k})),$$

ceci est fait au lemme 2.5.2 ;

$$\text{Var}(T_n(u_1, \dots, u_d)) = O\left(\frac{c(u_1, \dots, u_d)}{n(\frac{|\pi|}{k})^{d-d_1/2}}\right), \text{ si } |\pi| \text{ est fixé;}$$

$$\text{Var}(T_n(u_1, \dots, u_d)) = O\left(\frac{c(u_1, \dots, u_d)}{n(\frac{|\pi|}{k})^d}\right), \text{ si } k \text{ est fixé.}$$

Ceci est fait au lemme 2.5.4. Il nous reste à montrer que les autres termes de la variance de g_k sont négligeables devant la variance de T_n , $n > 0$.

$$\begin{aligned} \text{Var}(g_k(u_1, \dots, u_d)) &= \text{Var}(T_n(u_1, \dots, u_d)) + \text{Var}((T_n - g_k)(u_1, \dots, u_d)) \\ &\quad + 2\text{cov}(T_n(u_1, \dots, u_d), (g_k - T_n)(u_1, \dots, u_d)). \end{aligned}$$

Nous avons

$$\text{Var}((T_n - g_k)(u_1, \dots, u_d)) \leq \|(T_n - g_k) - E(T_n - g_k)\|^2$$

qui est négligeable devant $\text{Var}((T_n)(u_1, \dots, u_d))$ grâce à la condition (K5). De plus,

$$\text{cov}(T_n(u_1, \dots, u_d), (g_k - T_n)(u_1, \dots, u_d)) \leq \frac{\sqrt{\text{Var}(T_n(u_1, \dots, u_d))}}{\sqrt{\text{Var}((T_n - g_k)(u_1, \dots, u_d))}}$$

qui est négligeable devant $\text{Var}((T_n)(u_1, \dots, u_d))$ grâce à la condition (K5). \square

En ce qui concerne la convergence presque sûre, elle est donnée par le théorème 2.3.4. Pour fixer le cadre par rapport aux estimateurs concurrents, notons que la vitesse de convergence uniforme presque sûre de l'estimateur de la densité par la méthode des noyaux est de $\sqrt{\log n / (nh^d)}$ [Hansen, 2008; Deheuvels et Mason, 1992], h étant la fenêtre. Nous montrons que les mêmes vitesses de convergence s'appliquent pour l'estimateur spline de la densité basé sur les pseudo-observations.

Théorème 2.3.4. *Supposons que la densité de copule est bornée inférieurement et supérieurement par des réels strictement positifs, et continue dans un voisinage de $(u_1, \dots, u_d) \in [0, 1]^d$. Supposons que les conditions suivantes sont satisfaites :*

$$(K1) \quad \lim_{n \rightarrow \infty} n\left(\frac{|\pi|}{k}\right)^d = \infty, \quad \lim_{n \rightarrow \infty} \frac{|\pi|}{k} = 0,$$

$$(K2) \lim_{n \rightarrow \infty} \left(\frac{|\pi|}{k}\right)^d (\epsilon)^{-1} \in (a, \infty], a > 0,$$

$$(K3) \lim_{n \rightarrow \infty} \log(h_n^{-d}) = o(nh_n^d),$$

$$(K4) \lim_{n \rightarrow \infty} \log(h_n^{-d})(\log \log n)^{-1} = \infty$$

Sous les conditions (K1), (K2), (K3), (K4), c bornée sur un voisinage de (u_1, \dots, u_d) , nous avons :

$$\limsup_{n \rightarrow \infty} \sqrt{\left(\frac{|\pi|}{k}\right)^d \frac{n}{\log n}} |g_k(u_1, \dots, u_d) - E(g_k(u_1, \dots, u_d))| < \infty.$$

En plus, $(g_k - c)(u_1, \dots, u_d)$ converge presque sûrement vers 0 quand n tend vers l'infini, $(u_1, \dots, u_d) \in [0, 1]^d$. La convergence est uniforme si la densité de la copule est continue en tout point.

DÉMONSTRATION. Nous avons

$$\begin{aligned} |g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d)| &\leq \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} \left[\frac{|\alpha'_n(A_{j_1 \dots j_d}) - \alpha'_n(A_{j_1 \dots j_d})|}{\sqrt{n}} \right. \\ &\quad \left. \prod_{i=1}^d M_{j_i-k, k}(u_i) \right], \\ &\leq B \frac{1}{\sqrt{n}} \left(\frac{|\pi|}{k}\right)^{-d} \\ &\quad \sup_{j_1, \dots, j_d \in \{1, \dots, k+m\}} |\alpha'_n(A_{j_1 \dots j_d}) - \alpha_n(A_{j_1 \dots j_d})|, \end{aligned}$$

$B > 0$, d'après la propriété de partition de l'unité des M -splines établie au lemme 2.5.1 que nous rappelons :

$$I_{[0,1]^d}(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} \prod_{l=1}^d (\xi_{j_l-k+1} - \xi_{j_l-k}) M_{j_l-k, k}(u_l).$$

Or, en utilisant le résultat du lemme 2.3.1 nous pouvons écrire

$$\begin{aligned} |\alpha'_n(A_{j_1 \dots j_d}) - \alpha_n(A_{j_1 \dots j_d})| &\leq |\alpha_n(A_{j_1 \dots j_d}^+) - \alpha_n(A_{j_1 \dots j_d}^-)| \\ &\quad + \sqrt{n}(V_C(A_{j_1 \dots j_d}^+) - V_C(A_{j_1 \dots j_d}^-)) \\ &\leq |\alpha_n(A_{j_1 \dots j_d}^+) - \alpha_n(A_{j_1 \dots j_d}^-)| \\ &\quad + \sqrt{n}(V_C(A_{j_1 \dots j_d}^+) - V_C(A_{j_1 \dots j_d}^-)) \\ &\leq 2B_1 \sqrt{2\epsilon^d \log(\epsilon^d)} + B_2 \sqrt{n}\epsilon \end{aligned}$$

avec

$$\begin{aligned} A_{j_1 \dots j_d}^+ &= \prod_{l=1}^d [\xi_{j_l-k} - \epsilon, \xi_{j_l-k+1} + \epsilon), \\ A_{j_1 \dots j_d}^- &= \prod_{l=1}^d [\xi_{j_l-k} + \epsilon, \xi_{j_l-k+1} - \epsilon), \end{aligned}$$

$j_1, \dots, j_d = 1, \dots, k + m$. Nous en déduisons que

$$\begin{aligned} |\alpha'_n(A_{j_1 \dots j_d}) - \alpha_n(A_{j_1 \dots j_d})| &\leq B_1 \sup_{t \in A_{j_1 \dots j_d}} [\omega_n(\epsilon, t) + \omega_n(-\epsilon, t)] \\ &\quad + \sqrt{n} V_C(A_{j_1 \dots j_d}^+ \setminus A_{j_1 \dots j_d}^-). \end{aligned}$$

$B_1 > 0$. Si la densité c est bornée sur $A_{j_1 \dots j_d}^+$, $j_1, \dots, j_d \in \{1, \dots, k + m\}$, nous pouvons trouver $B_2 > 0$ tel que

$$\begin{aligned} |\alpha'_n(A_{j_1 \dots j_d}) - \alpha_n(A_{j_1 \dots j_d})| &\leq B_1 \sup_{t \in A_{j_1 \dots j_d}} [\omega_n(\epsilon, t) + \omega_n(-\epsilon, t)] \\ &\quad + B_2 \sqrt{n} \epsilon \left(\frac{|\pi|}{k}\right)^{d-1}, \end{aligned}$$

$B_2 > 0$. Finalement, nous avons

$$\begin{aligned} \sqrt{\left(\frac{|\pi|}{k}\right)^d \frac{n}{\log n}} |g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d)| &\leq 2BB_1 \sqrt{\frac{2\epsilon^d |\log(\epsilon^d)|}{\log(n) \left(\frac{|\pi|}{k}\right)^d}} \\ &\quad + BB_2 \epsilon \sqrt{\frac{n}{\log n} \left(\frac{|\pi|}{k}\right)^{d/2-1}}, \end{aligned}$$

et on en déduit que

$$\limsup_{n \rightarrow \infty} \sqrt{\left(\frac{|\pi|}{k}\right)^d \frac{n}{\log n}} |g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d)| < \infty,$$

et que

$$\limsup_{n \rightarrow \infty} \sqrt{\left(\frac{|\pi|}{k}\right)^d \frac{n}{\log n}} |E(g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d))| < \infty.$$

On refait le même travail pour $T_n(u_1, \dots, u_d)$; en remarquant que $E(T_n) = f_n$, nous avons

$$\begin{aligned} |T_n(u_1, \dots, u_d) - E(T_n(u_1, \dots, u_d))| &= \frac{1}{\sqrt{n}} [\sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} \alpha_n(A_{j_1 \dots j_d})], \\ &\leq B' \frac{1}{\sqrt{n}} \left(\frac{|\pi|}{k}\right)^{-d} \sup_{j_1, \dots, j_d \in \{1, \dots, k+m\}} |\alpha_n(A_{j_1 \dots j_d})|, \end{aligned}$$

$B' > 0$. On en déduit que

$$\sqrt{\left(\frac{|\pi|}{k}\right)^d \frac{n}{\log n}} |T_n(u_1, \dots, u_d) - E(T_n(u_1, \dots, u_d))| \leq 2B'B_3 \sqrt{\frac{2|\log((\frac{|\pi|}{k})^d)|}{\log(n)}}.$$

En utilisant l'inégalité triangulaire

$$\begin{aligned} |g_k(u_1, \dots, u_d) - E(g_k(u_1, \dots, u_d))| &\leq |g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d)| \\ &\quad + |T_n(u_1, \dots, u_d) - E(T_n(u_1, \dots, u_d))| \\ &\quad + |E(g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d))|, \end{aligned}$$

on arrive au résultat. Pour montrer que $(g_k - c)(u_1, \dots, u_d)$ converge presque sûrement vers 0, on peut utiliser l'inégalité

$$\begin{aligned} |g_k(u_1, \dots, u_d) - c(u_1, \dots, u_d)| &\leq |g_k(u_1, \dots, u_d) - T_n(u_1, \dots, u_d)| \\ &+ |T_n(u_1, \dots, u_d) - E(T_n(u_1, \dots, u_d))| \\ &+ |E(T_n(u_1, \dots, u_d)) - c(u_1, \dots, u_d)|. \end{aligned}$$

Les deux premiers termes de l'inégalité convergent. La convergence du troisième terme est donnée par le lemme 2.5.2. \square

Pour $\frac{|\pi|}{k} = O((\log n/n)^{\frac{1}{d+2}})$, k fixé, nous atteignons la vitesse optimale de la convergence uniforme presque sûre pour une densité (bornée inférieurement et supérieurement par des réels positifs) dont les dérivées partielles d'ordre un sont bornées.

2.4. DISCUSSION

L'estimateur spline de la densité de copule est une densité de copule. Il réalise un meilleur compromis biais-variance que l'estimateur de la densité de la copule de Bernstein. Le biais est de l'ordre h pour l'estimateur spline de la densité de copule tandis qu'il est de l'ordre h^2 sur $(0, 1)^2$ pour l'estimateur par la méthode des noyaux, h représentant la fenêtre (pour l'estimateur spline h équivaut à π/k , π distance maximale entre deux nœuds distincts consécutifs, k l'ordre des splines). Cependant, cet avantage de l'estimateur par noyau n'est obtenu qu'au prix de l'hypothèse que la densité de la copule à estimer est deux fois dérivable, hypothèse qui n'est pas nécessaire dans notre cadre de travail. Les résultats présentés dans ce travail sont avant tout d'ordre asymptotique : nous avons constaté que pour n (nombre d'observations) petit, l'estimateur de la densité n'est pas une très bonne approximation de la densité de copule aux bords (surtout si la densité est infinie aux bords). Sur les simulations (figures 2-1 à 2-6), nous présentons les estimations de la copule et de la densité de copule dans le cas des familles de copules de Clayton, Frank, Gumbel, et des copules gaussiennes. Il nous a fallu prendre au moins $n = 10^6$ pour que la variance de l'estimateur respecte l'ordre de grandeur spécifié dans cet article.

2.5. RÉSULTATS INTERMÉDIAIRES ET PREUVES DES THÉORÈMES

Pour la clarté de l'exposition des résultats, rappelons les quantités

$$T_n(u_1, \dots, u_d) = \frac{1}{n} \sum_{i=1}^n \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} I_{A_{j_1 \dots j_d}}(F_1(X_{1i}), \dots, F_d(X_{di})) \prod_{l=1}^d M_{j_l-k, k}(u_l),$$

et

$$f_n(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} V_C(A_{j_1 \dots j_d}) \prod_{l=1}^d M_{j_l-k, k}(u_l),$$

$(u_1, \dots, u_d) \in [0, 1]^d$, V_C est la fonction volume associée à la copule C , $A_{j_1 \dots j_d} = \prod_{l=1}^d [\xi_{j_l-k}, \xi_{j_l-k+1})$, $j_i = 1, \dots, k+m$, $i = 1, \dots, d$. Nous avons $E(T_n) = f_n$.

Lemme 2.5.1. *Résultat intermédiaire pour le théorème 2.3.4*

Soit C_0 la copule indépendante vérifiant

$$C_0(u_1, \dots, u_d) = u_1 \cdots u_d$$

$(u_1, \dots, u_d) \in [0, 1]^d$. Nous avons :

$$C_0(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} \prod_{l=1}^d (\xi_{j_l-k+1} - \xi_{j_l-k}) G_{j_l k}(u_l), \quad (2.5.1)$$

$$I_{[0,1]^d}(u_1, \dots, u_d) = \sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} \prod_{l=1}^d (\xi_{j_l-k+1} - \xi_{j_l-k}) M_{j_l-k, k}(u_l). \quad (2.5.2)$$

DÉMONSTRATION. Nous devons démontrer que

$$C_0(u_1, \dots, u_d) = \prod_{l=1}^d \sum_{j_l=1}^{k+m} (\xi_{j_l-k+1} - \xi_{j_l-k}) G_{j_l k}(u_l).$$

Or, les équations

$$\sum_{j_l=1}^{k+m} (\xi_{j_l-k+1} - \xi_{j_l-k}) G_{j_l k}(u_l) = u_l,$$

$l = 1, \dots, d$ ne sont qu'une reformulation du lemme A.2.1. Ils permettent d'aboutir à l'égalité (2.5.1). L'équation (2.5.2) n'est qu'une reformulation de (2.5.1) pour la fonction de densité. \square

Lemme 2.5.2. *Résultat intermédiaire pour le théorème 2.3.4*

Si c la densité de la copule est continue dans un voisinage de $(u_1, \dots, u_d) \in [0, 1]^d$, nous avons les inégalités suivantes :

$$\begin{aligned} |f_n(u_1, \dots, u_d) - c(u_1, \dots, u_d)| &\leq d \frac{|\pi|}{kh} \omega_1(c, h, \dots, h) \\ &+ (1 + \frac{1}{h^2} (1 + \sqrt{\frac{k}{12}})^2 \frac{|\pi_k|^2}{k^2}) \omega_2(c, h, \dots, h). \end{aligned}$$

avec $|\pi| = \max_j (x_{j-k} - x_j)$. Si c est continue en (u_1, \dots, u_d) , $f_n(u_1, \dots, u_d)$ converge vers $c(u_1, \dots, u_d)$.

DÉMONSTRATION. Nous avons

$$|(f_n - c)(u_1, \dots, u_d)| = |(Q_\pi c - c)(u_1, \dots, u_d)|,$$

Q_π est l'opérateur lissage par splines. Nous avons par conséquent

$$|f_n(u_1, \dots, u_d) - c(u_1, \dots, u_d)| \leq d \frac{|\pi_k|}{kh} \omega_1(c, h, \dots, h) \\ + (1 + \frac{1}{h^2} (1 + \sqrt{\frac{k}{12}})^2 \frac{|\pi_k|^2}{k^2}) \omega_2(c, h, \dots, h).$$

□

Lemme 2.5.3. Soit μ_0 la mesure de Lebesgue sur $[0, 1]^d$, pour k assez grand,

$$\sum_{j_1=1}^{k+m} \cdots \sum_{j_d=1}^{k+m} \mu_0(A_{j_1 \dots j_d})^2 \prod_{l=1}^d M_{j_l-k, k}^2(u_l) = O((\frac{|\pi_k|}{k})^{d_1/2})$$

avec d_1 le nombre d'éléments parmi u_1, \dots, u_d différents de 0 et de 1.

DÉMONSTRATION. Pour l fixé, soit $u_l \in [0, 1]$

- Si $u_l \in \{0, 1\}$, supposons pour fixer les idées $u_l = 0$. Nous avons pour $j_l \neq j'_l$

$$M_{j_l-k, k}(u_l) M_{j'_l-k, k}(u_l) = 0.$$

En effet, d'après l'équation (A.1.2) et le lemme 2.2.1, nous avons

$$\sum_{j=1}^{m+k} (\frac{1}{k-1} \sum_{l=1}^{k-1} x_{j-k+l}) N_{j-k, k}(0) = 0.$$

Comme les M-splines sont des B-splines à une constante près, les $M_{j_l-k, k}(0)$ sont tous nuls à l'exception de $M_{1-k, k}(0)$.

- Si $u_l \notin \{0, 1\}$, nous avons pour $j_l \neq j'_l$, nous utilisons l'approximation des splines par la densité d'une loi normale. Posons

$$Z^k = \delta_{j+1-k} U_{(k:k)} + \cdots + \delta_j U_{(1:k)},$$

avec

$$\delta_{j+1-k} = x_{j+1-k} - x_{j-k}, \dots, \delta_j = x_{j+1} - x_j,$$

, $i = 1, \dots, m$; $U_{(l:k)}$ est la l ème statistique d'ordre parmi U_1, \dots, U_k , variables aléatoires de loi uniforme sur $[0, 1]$, $l = 1, \dots, k$, $j = 1-k, \dots, m$. La fonction G_{jk} , si elle est continue, est la fonction de répartition de $x_{j-i} + Z^k$, $j = 1, \dots, m+i$, $i = 1, \dots, k$. Ce résultat est la conséquence directe du fait que la densité de $x_{j-k} + Z^k$ est la M-spline dont les nœuds sont x_{j-k}, \dots, x_j , (voir [Adell et Sanguesa, 2005; Karlin, Micchelli et Rinott, 1986]). En utilisant le fait que les espacements d'ordre 1 entre statistiques d'ordre sont asymptotiquement indépendants et de loi bêta, nous avons

$$E(Z^k) = \frac{\sum_{l=0}^k x_{j-k+l+1}}{k+1} = O(u_l), \\ \tau^2 = V(Z^k) = O(\frac{k}{(k+1)^2(k+2)} \sum_{l=0}^k x_{j-k+l+1}^2) = O(u_l^2 \frac{\pi}{k}).$$

$$\begin{aligned} \sum_{j_l=1}^{k+m} \left(\frac{\pi}{k}\right)^2 M_{j_l-k, k}^2(u_l) &\approx \\ \int \left(\frac{\pi}{k}\right)^2 \left(\frac{1}{\sqrt{2\pi\tau}} \exp -\frac{(u_l - \frac{\pi}{k}z)^2}{2\tau^2}\right)^2 dz &= O\left(\frac{\pi}{k}\tau^{-1}\right), \\ &= O\left(\sqrt{\frac{\pi}{k}}\right). \end{aligned}$$

□

Lemme 2.5.4.

$$\text{Var}(T_n(u_1, \dots, u_d)) = \frac{1}{n} \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} \mu(A_{j_1 \dots j_d})(1 - \mu(A_{j_1 \dots j_d})) \prod_{l=1}^d M_{j_l-k, k}^2(u_l).$$

Si la densité de la copule c est continue au voisinage de (u_1, \dots, u_d) , $(u_1, \dots, u_d) \in [0, 1]^d$, et n assez grand, nous avons

$$\text{Var}(T_n(u_1, \dots, u_d)) = O\left(\frac{c(u_1, \dots, u_d)}{n\left(\frac{\pi}{k}\right)^{d-d_1/2}}\right)$$

avec d_1 le nombre d'éléments parmi u_1, \dots, u_d différents de 0 et de 1.

DÉMONSTRATION. Nous avons

$$\text{Var}(T_n(u_1, \dots, u_d)) = \frac{1}{n} \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} \text{Var}(I_{A_{j_1 \dots j_d}}(F_1(X_{11}), \dots, F_d(X_{d1}))) \prod_{l=1}^d M_{j_l-k, k}^2(u_l),$$

d'où

$$\text{Var}(T_n(u_1, \dots, u_d)) = \frac{1}{n} \sum_{j_1=1}^{k+m} \dots \sum_{j_d=1}^{k+m} V_C(A_{j_1 \dots j_d})(1 - \mu(A_{j_1 \dots j_d})) \prod_{l=1}^d M_{j_l-k, k}^2(u_l),$$

V_C est la fonction volume associée à C . Si la densité de la copule c est continue au voisinage de (u_1, \dots, u_d) , $(u_1, \dots, u_d) \in [0, 1]^d$, et n assez grand, nous avons

$$V_C(A_{j_1 \dots j_d}) = O(c(u_1, \dots, u_d)\mu_0(A_{j_1 \dots j_d})),$$

μ_0 étant la mesure de Lebesgue sur $[0, 1]^d$. Enfin, le fait que

$$\mu_0(A_{j_1 \dots j_d}) = O\left(\left(\frac{\pi}{k}\right)^d\right),$$

et le lemme 2.5.3 permettent de conclure. □

2.6. CONCLUSION

Nous proposons des estimateurs spline de la copule et de la densité de copule basés sur des pseudo-observations. Ce sont des généralisations des estimateurs copules de Bernstein empiriques. L'estimateur de la densité ainsi construit ne souffre pas de biais aux frontières et atteint la vitesse de convergence optimale pour une densité dérivable une fois et bornée inférieurement et supérieurement par des nombres réels positifs. Cet estimateur pourrait être utilisé dans le cadre d'une analyse multirésolution de la copule de Bernstein par insertion de nœuds.

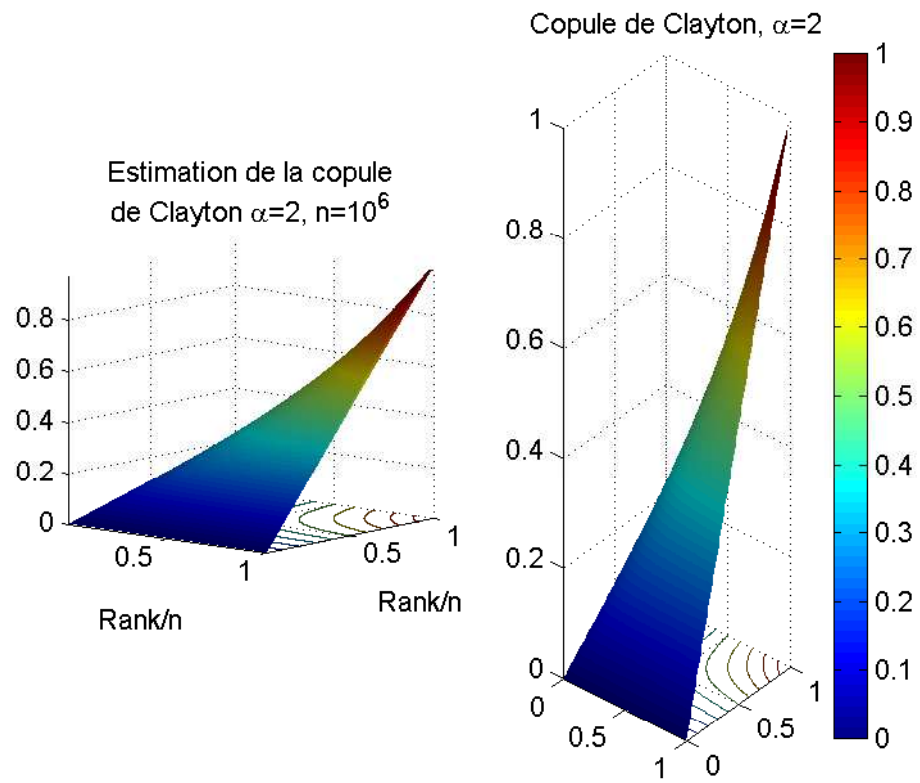


FIGURE 2.1. Copule de Clayton

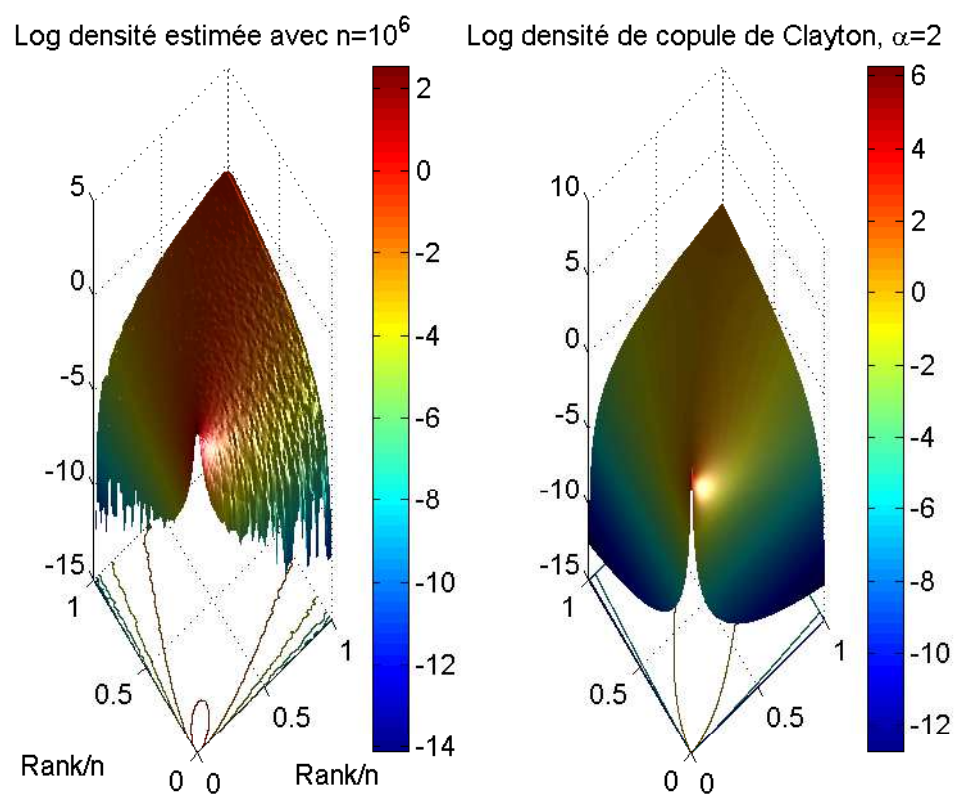


FIGURE 2.2. Log densité de la copule de Clayton

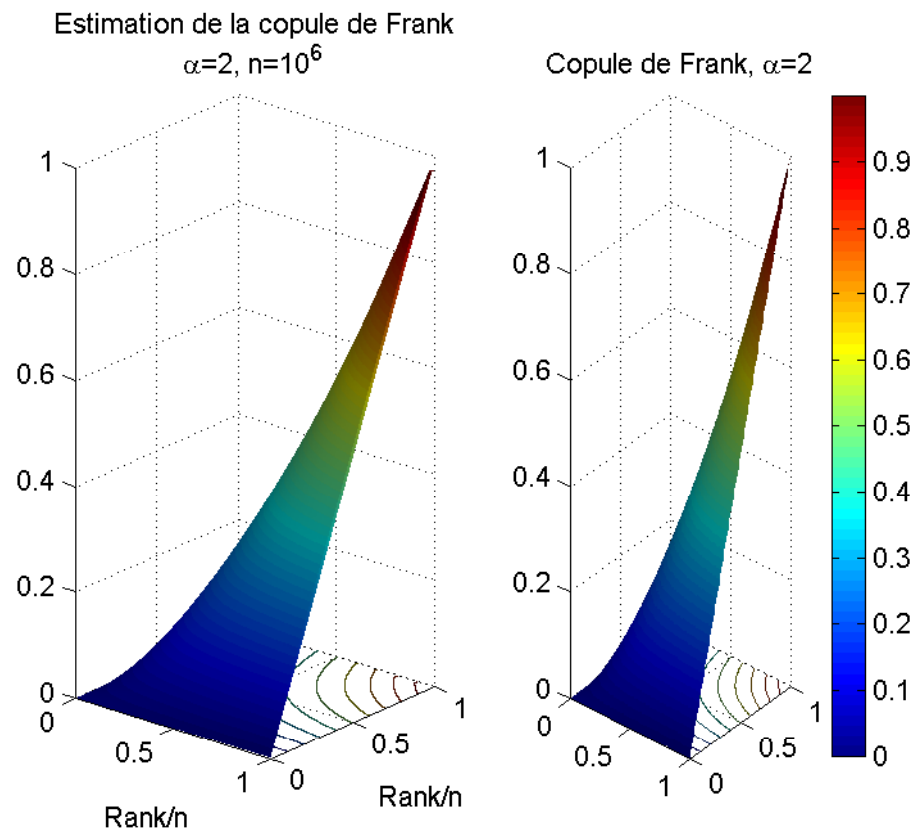


FIGURE 2.3. Copule de Frank

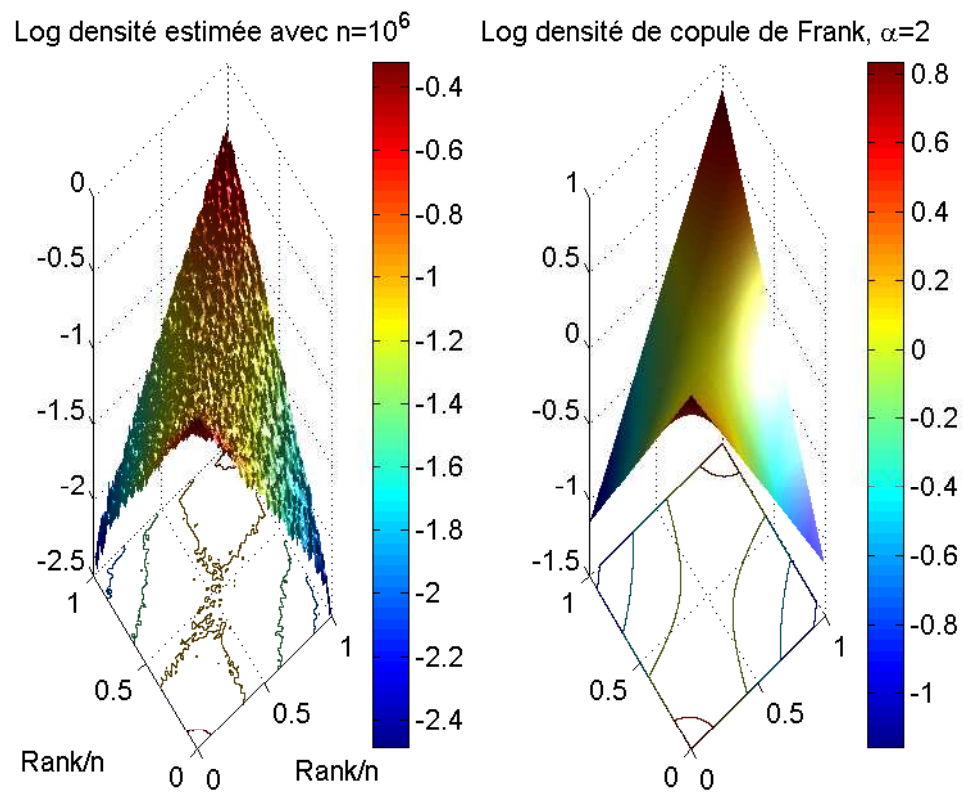


FIGURE 2.4. Log densité de la copule de Frank

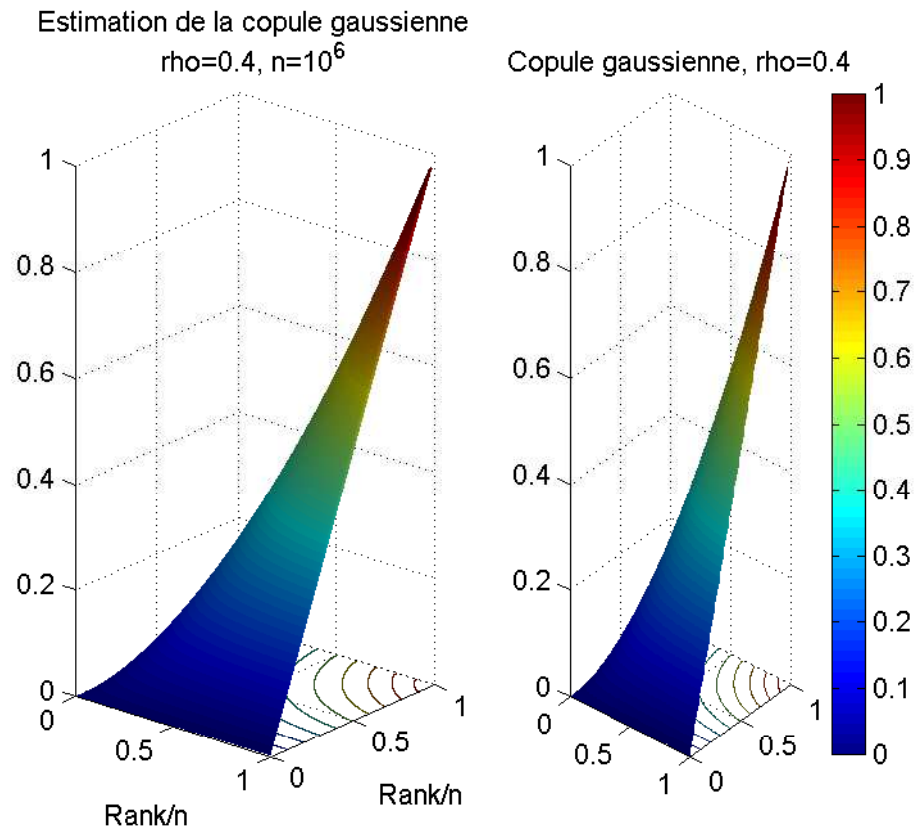


FIGURE 2.5. Copule gaussienne

Log densité estimée avec $n=10^6$ Log densité de copule gaussienne, $\rho=0.4$

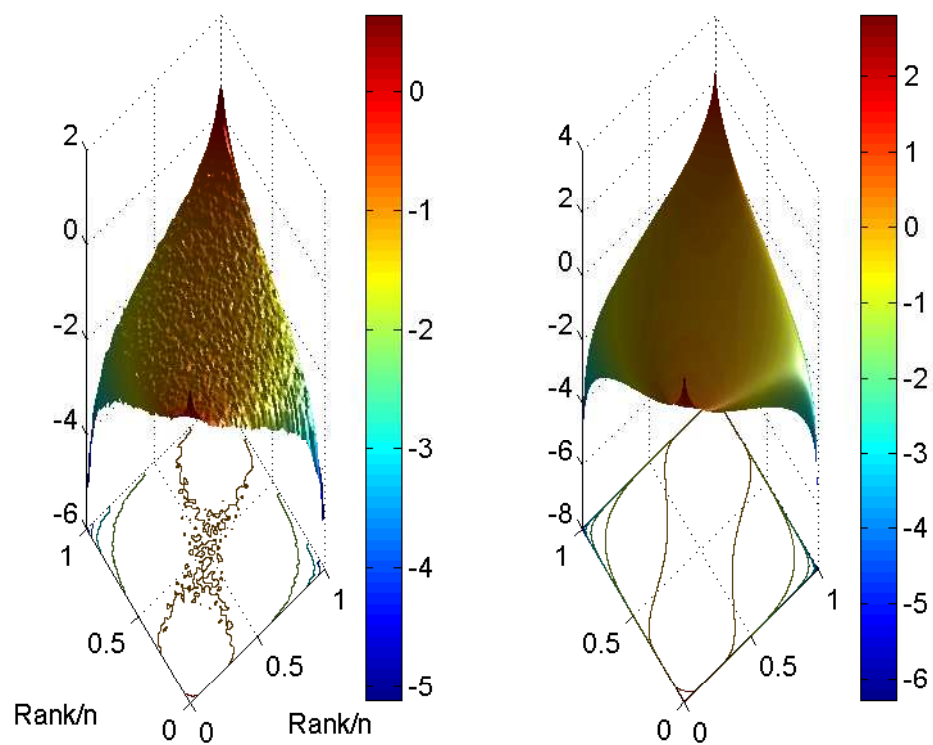


FIGURE 2.6. Log densité de la copule gaussienne

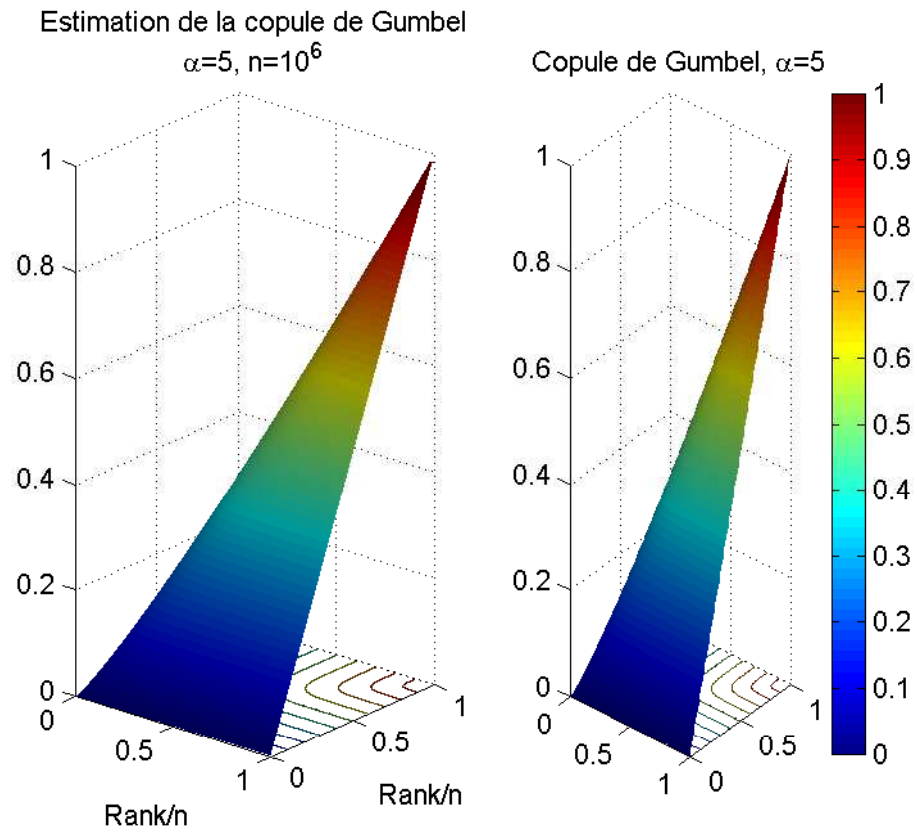


FIGURE 2.7. Copule de Gumbel

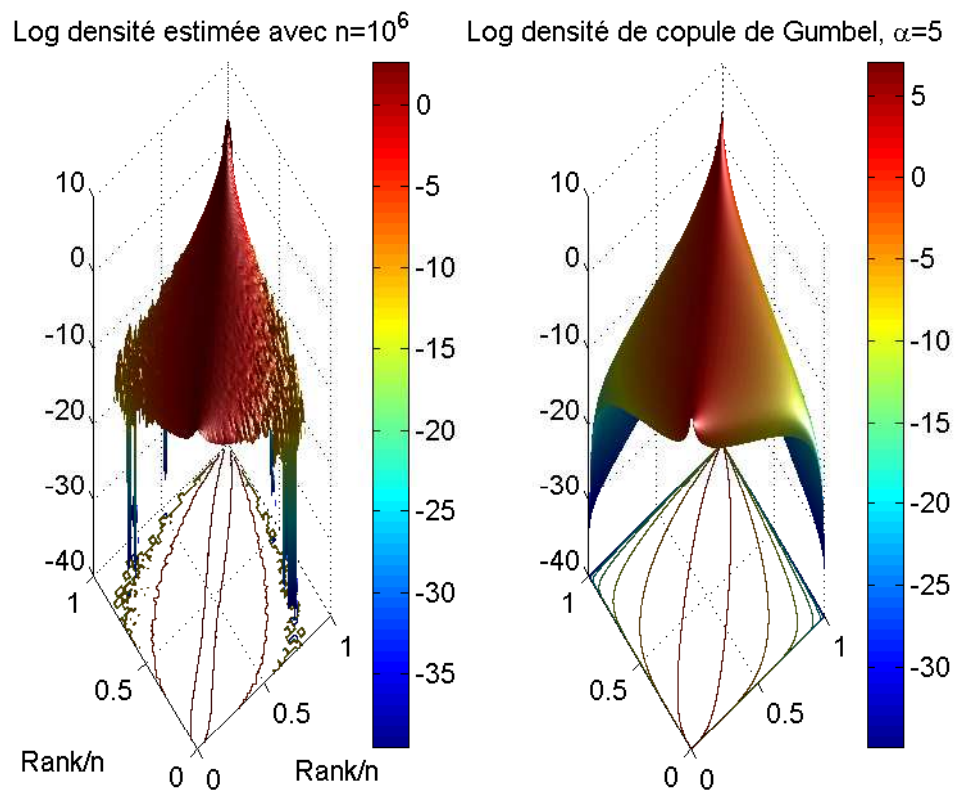


FIGURE 2.8. Log densité de la copule de Gumbel

Chapitre 3

CARACTÉRISATION DES PROLONGEMENTS D'UNE SOUS-COPULE EN DIMENSION 2

Résumé. *L'objectif du chapitre est de caractériser l'ensemble des copules qui prolongent une sous-copule donnée en dimension 2. Ce sujet a été traité par le passé par [Nešlehová, 2007], [Faugeras, 2012], [de Amo et coll., 2012]. Cependant, il reste des zones d'ombre que ce chapitre cherche à combler.*

3.1. INTRODUCTION

Soit F une fonction de répartition en dimension 2 avec F_1, F_2 les fonctions de répartition marginales. Nous disons que D est une sous-copule générée par F si S_i est l'espace image de F_i , $i = 1, 2$, D est définie sur $S_1 \times S_2$ par la relation $D(F_1(x_1), F_2(x_2)) = F(x_1, x_2)$ pour tout $(x_1, x_2) \in \mathbb{R}^2$. Une sous-copule D est générée par une fonction de répartition F si et seulement si D est une sous-copule et $\inf\{t \in S_i : t \geq u_i\} \in S_i$ pour tout $u_i \in [0, 1]$, $i = 1, 2$, $S_1 \times S_2$ étant le domaine de D .

Soit \mathcal{C} l'espace des copules et \mathcal{D} l'espace des sous-copules. Nous disons que $C \in \mathcal{C}$ est un prolongement de $D \in \mathcal{D}$ si la restriction de C sur $S_1 \times S_2$ est donnée par D . Soit $\mathcal{C}(D)$ l'espace des prolongements de D , $D \in \mathcal{D}$. Il existe un lien entre les prolongements d'une sous-copule générée par une fonction de répartition et le théorème de Sklar. Le théorème de Sklar dit que si F est une fonction de répartition alors il existe une copule C telle que

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)), (x_1, x_2) \in \mathbb{R}^2,$$

C est définie de façon unique sur $\{(F_1(x_1), F_2(x_2)) : (x_1, x_2) \in \mathbb{R}^2\}$. L'objectif de ce chapitre est de donner une formule explicite de tous les prolongements possibles dans le cadre le plus général possible.

Si D est générée par une fonction de répartition, par le théorème de Sklar, nous pouvons montrer que $\mathcal{C}(D) \neq \emptyset$. La toute première preuve du théorème de Sklar peut être trouvée dans [Schweizer et Sklar, 1974]. D'autres versions et prolongements peuvent être trouvés dans [Carley et Taylor, 2002], [Denuit et Lambert, 2005], [Rüschendorf, 2009]. Un autre type de prolongement basé sur la transformation de [Szczesny, 1991] a été proposé par [Nešlehová, 2007]. Se basant sur une autre version plus générale de la transformation de [Szczesny, 1991] donnée par [Shorack, 2000], [Faugeras, 2012] a proposé une caractérisation des prolongements des sous-copules discrètes.

Nous dirons que C^{\max} est le prolongement maximal de la sous-copule D si $C^{\max} \in \mathcal{C}(D)$ et $C^{\max} \geq C$ pour tout $C \in \mathcal{C}(D)$. De la même façon, C^{\min} est le prolongement minimal de la sous-copule D si $C^{\min} \in \mathcal{C}(D)$ et $C^{\min} \leq C$ pour tout $C \in \mathcal{C}(D)$. Le plus souvent, les prolongements d'une sous-copule ne sont pas uniques mais, le prolongement maximal et le prolongement minimal existent. Ce résultat a été prouvé par [Carley, 2002] lorsque le domaine de D est de cardinalité finie. Une généralisation au cas où le domaine est dénombrable a été faite par [Genest et Nešlehová, 2007]. À part la caractérisation de tous les prolongements de la sous-copule D , un autre objectif du chapitre est de montrer que les prolongements maximal et minimal existent dans le cadre le plus général possible. Apparemment, nos objectifs ont été atteints par [de Amo et coll., 2012]. Cependant, la caractérisation de [de Amo et coll., 2012] implique des ensembles P_1 et P_2 appelés ensembles de singletons (Section 3). Il est implicite dans la construction que P_1 et P_2 sont de mesure de Lebesgue nulle. Nous pouvons trouver des contre-exemples en utilisant l'ensemble de Smith-Volterra-Cantor par exemple, voir [Aliprantis, 1998]. Nous proposons donc une construction dans le cadre le plus général possible au théorème 3.2.1.

Dans la section 2, nous donnons une construction caractérisant tous les prolongements d'une sous-copule D (théorème 3.2.1). L'idée de base s'appuie sur la règle des probabilités totales. Dans les section 3 et 4, nous développons une expression pour C^{\max} et C^{\min} . Notons que C^{\max} et C^{\min} sont liées par la relation

$$C^{\min}(u, v) = v - C^{\max}(1 - u, v), u, v \in [0, 1].$$

À la section 5, nous discutons des limitations de l'approche proposée par [de Amo et coll., 2012].

Dans ce chapitre, l'intérieur d'un ensemble E est noté \mathring{E} . La fermeture d'un ensemble E est notée \overline{E} . L'inverse généralisé de la fonction de répartition F sur $[0, 1]$ est noté $F^{(-1)}$ et est donné par $F^{(-1)}(u) = \inf\{t \in [0, 1] : F(t) \geq u\}$, $u \in [0, 1]$.

3.2. LES PROLONGEMENTS

Soit D une sous-copule sur $S_1 \times S_2$. Nous voulons trouver l'ensemble des copules prolongeant D . Le lemme 3.2.1 et lemme 3.2.2 donnent des prolongements naturels. Ces prolongements sont uniques. Après le lemme 3.2.1, nous donnons une partition de $[0, 1]^2$. Cette partition est l'élément clé du chapitre. Le lemme 3.2.3 donne l'application de la règle des probabilités totales appliquée aux copules. Le résultat général est donné au théorème 3.2.1.

Lemme 3.2.1. *Soit D une sous-copule définie sur $S_1 \times S_2$ alors D admet un prolongement en une sous-copule \overline{D} sur $\overline{S_1} \times \overline{S_2}$. Nous avons $C^{\max} \in \mathcal{C}(D)$ si et seulement si $C^{\max} \in \mathcal{C}(\overline{D})$.*

DÉMONSTRATION. Si D est une sous-copule sur $S_1 \times S_2$ alors D est Lipschitz-continue de constante de Lipschitz 1. En d'autres mots,

$$|D(u, v) - D(u', v')| \leq |u - u'| + |v - v'|, (u, v), (u', v') \in S_1 \times S_2.$$

Si $(u, v) \in \overline{S_1} \times \overline{S_2}$, alors nous pouvons trouver une suite $\{(u_n, v_n)\}_{n=1}^{\infty}$ dans $S_1 \times S_2$ convergeant vers (u, v) . Définissons

$$\overline{D}(u, v) = \lim_{n \rightarrow \infty} D(u_n, v_n).$$

La fonction \overline{D} est bien définie (ne dépend pas de la suite utilisée dans la définition) parce que D est Lipschitz-continue de constante de Lipschitz 1. Il est facile de vérifier que \overline{D} est une sous-copule et que C est une copule prolongeant D si et seulement si C est une copule prolongeant \overline{D} . \square

Trouvons une partition de $[0, 1]^2$ basée sur $\overline{S_1} \times \overline{S_2}$. Comme $\overline{S_k}$ est fermée, et que $\{0, 1\} \subset \overline{S_k}$, $k = 1, 2$, nous pouvons écrire

$$[0, 1] = \bigcup_{l \in J_k} V_{kl}, V_{k0} = \overline{S_k}, V_{kl} = (a_{kl}, b_{kl}), a_{kl} < b_{kl}, 0 < l \in J_k \subset \mathbb{N}, k = 1, 2.$$

Supposons que $\{0\} \in J_k$, soit $J_k^* = J_k \setminus \{0\}$, $k = 1, 2$. La partition est construite de la façon suivante. Nous prenons

$$E_{kl} = (a_{kl}, b_{kl}], l \in J_k^*, E_{k0} = [0, 1] \setminus \bigcup_{l \in J_k^*} E_{kl}, k = 1, 2$$

$$[0, 1]^2 = \bigcup_{(i,j) \in J_1 \times J_2} E_{1i} \times E_{1j}.$$

Considérons les classes d'ensembles \mathcal{A}_k , $k = 1, 2$, données par

$$\mathcal{A}_k = \{(a, b] : a \leq b, a, b \in \overline{S_k}\} \cup \{[0, b] : b \in \overline{S_k}\}, k = 1, 2.$$

Posons

$$\mathcal{A} = \{A_1 \times A_2 : A_k \in \mathcal{A}_k, k = 1, 2\}.$$

Il est aisé de voir que \mathcal{A} est une semi-algèbre. Pour $A_k \in \mathcal{A}_k$, un intervalle ayant comme extrémités a_k, b_k , $a_k \leq b_k$, $k = 1, 2$, nous considérons la fonction V_D définie par

$$V_D(A_1 \times A_2) = \overline{D}(a_1, b_1) - \overline{D}(a_2, b_1) - \overline{D}(a_1, b_2) + \overline{D}(a_2, b_2).$$

Lemme 3.2.2. *La fonction V_D définie sur \mathcal{A} admet un prolongement unique en une mesure de probabilité $P_{\overline{D}}$ sur $\sigma(\mathcal{A})$.*

La preuve du lemme 3.2.2 peut être trouvée en annexe. Soit \mathcal{H} l'espace des fonctions de répartition continues sur \mathbb{R} . Considérons $\{A_{1i} \times A_{2j} : (i, j) \in J_1 \times J_2\}$ une partition de $[0, 1]^2$, $J_k \subset \mathbb{N}$, $k = 1, 2$.

Lemme 3.2.3. (1) *Nous avons $C \in \mathcal{C}$ si et seulement si*

$$C(u, v) = \sum_{\{(i,j) \in J_1 \times J_2 : p_{ij} > 0\}} p_{ij} C_{ij}(H_{1ij}(u), H_{2ij}(v)), u, v \in [0, 1] \quad (3.2.1)$$

pour $p_{ij} \in [0, 1]$, $(i, j) \in J_1 \times J_2$ avec $C_{ij} \in \mathcal{C}$, $H_{kij} \in \mathcal{H}$, $(i, j) \in J_1 \times J_2$, $k = 1, 2$

$$\sum_{(i,j) \in J_1 \times J_2} p_{ij} = 1,$$

H_{1ij}, H_{2ij} sont des fonctions de répartition continues concentrées respectivement sur A_{1i} et A_{2j} , $(i, j) \in J_1 \times J_2$, $p_{ij} > 0$

$$\sum_{\{(i,j) \in J_1 \times J_2 : p_{ij} > 0\}} p_{ij} H_{1ij}(u) = P(U \in [0, u] \cap A_{1i}), u \in [0, 1], \quad (3.2.2)$$

$U \sim \mathcal{U}[0, 1]$ si

$$i \in J_1, \sum_{j \in J_2} p_{ij} > 0,$$

$$\sum_{\{(i,j) \in J_1 \times J_2 : p_{ij} > 0\}} p_{ij} H_{2ij}(v) = P(U \in [0, v] \cap A_{2j}), v \in [0, 1], \quad (3.2.3)$$

si

$$j \in J_2, \sum_{i \in J_1} p_{ij} > 0.$$

(2) *Supposons que C est donné par l'équation 3.2.1. Soit P_C la mesure de probabilité associée avec C et*

$\mathcal{M} = \{(i, j) \in J_1 \times J_2 : p_{ij} > 0\}$. Pour tout $u, v \in [0, 1]$, nous obtenons

$$p_{ij} = P_C(A_{1i} \times A_{2j}), (i, j) \in J_1 \times J_2$$

$$H_{1ij}(u) = P_C([0, u] \times [0, 1] | A_{1i} \times A_{2j}), (i, j) \in \mathcal{M}$$

$$H_{2ij}(v) = P_C([0, 1] \times [0, v] | A_{1i} \times A_{2j}), (i, j) \in \mathcal{M}$$

H_{1ij}, H_{2ij} sont Lipschitz-continues de constante de Lipschitz p_{ij}^{-1} , $(i, j) \in \mathcal{M}$.

$$C_{ij}(u, v) = P_C((-\infty, H_{1ij}^{(-1)}(u)] \times (-\infty, H_{2ij}^{(-1)}(v)] | A_{1i} \times A_{2j}), (i, j) \in \mathcal{M}.$$

(3) Soit P_C la mesure de probabilité associée à C , $C \in \mathcal{C}$. Considérons les fonctions de répartition $H_{1i\cdot}$, $H_{2\cdot j}$ données par

$$H_{1i\cdot}(u) = P(U \leq u | U \in A_{1i}) \text{ pour } P(U \in A_{1i}) > 0, U \sim \mathcal{U}(0, 1), i \in J_1$$

$$H_{2\cdot j}(v) = P(U \leq v | U \in A_{2j}) \text{ pour } P(U \in A_{2j}) > 0, U \sim \mathcal{U}(0, 1), j \in J_2.$$

Les équations (3.2.2) et (3.2.3) sont satisfaites si et seulement si

$$H_{1ij} = \Phi_{1ij} \circ \Phi_{1i\cdot}^{(-1)} \circ H_{1i\cdot}$$

avec

$$\Phi_{1i\cdot} = \frac{\sum_{\{k:(i,k) \in \mathcal{M}\}} p_{ik} \Phi_{1ik}}{\sum_{\{k:(i,k) \in \mathcal{M}\}} p_{ik}},$$

$$(i, j) \in \mathcal{M},$$

$$H_{2ij} = \Phi_{2ij} \circ \Phi_{2\cdot j}^{(-1)} \circ H_{2\cdot j}$$

avec

$$\Phi_{2\cdot j} = \frac{\sum_{\{k:(k,j) \in \mathcal{M}\}} p_{kj} \Phi_{2kj}}{\sum_{\{k:(k,j) \in \mathcal{M}\}} p_{kj}},$$

$$(i, j) \in \mathcal{M}, \text{ pour } \Phi_{1ij}, \Phi_{2ij} \in \mathcal{H}, (i, j) \in \mathcal{M}.$$

DÉMONSTRATION. (parties a et b) Il est facile de voir que si C satisfait l'équation (3.2.1) avec H_{1ij} et H_{2ij} des fonctions de répartition concentrées sur A_{1i} , A_{2j} , $p_{i,j} > 0$, $(i, j) \in \mathcal{M}$ et $\sum_{(i,j) \in J_1 \times J_2} p_{ij} = 1$ alors C est une fonction de répartition

concentrée sur $[0, 1]$. Soit P_C la mesure de probabilité associée à C . À partir des équations (3.2.2) et (3.2.3), nous obtenons

$$P_C([0, u] \times [0, 1] \cap A_{1i} \times [0, 1]) = P(U \leq u, U \in A_{1i}),$$

pour $P(U \in A_{1i}) > 0, i \in J_1$;

$$P_C([0, 1] \times [0, v] \cap [0, 1] \times A_{2j}) = P(U \leq v, U \in A_{2j}),$$

pour $P(U \in A_{2j}) > 0, j \in J_2$. Ceci implique que C a des marges uniformes, donc $C \in \mathcal{C}$.

Supposons que $C \in \mathcal{C}$. Soit P_C la mesure de probabilité associée à C . Soit

$$p_{ij} = P_C(A_{1ij} \times A_{2ij}), (i, j) \in \mathcal{M}.$$

Nous pouvons écrire

$$C(u, v) = \sum_{(i,j) \in J_1 \times J_2 : p_{ij} > 0} p_{ij} P_C([0, u] \times [0, v] | A_{1ij} \times A_{2ij}), u, v \in [0, 1].$$

Soit $P_C(\cdot | A_{1ij} \times A_{2ij})$ la probabilité conditionnelle induite par P_C , $(i, j) \in \mathcal{M}$. Soient H_{1ij} et H_{2ij} les fonctions de répartition marginales associées à $P_C(\cdot | A_{1ij} \times A_{2ij})$, $(i, j) \in \mathcal{M}$. Si $0 \leq u_0 < u_1 \leq 1$ alors

$$\begin{aligned} p_{ij}(H_{1ij}(u_1) - H_{1ij}(u_0)) &\leq \sum_{k: (i,k) \in \mathcal{M}} p_{ik}(H_{1ij}(u_1) - H_{1ij}(u_0)), \\ &= P_C((u_0, u_1] \cap A_{1ij} \times [0, 1]) \\ &\leq u_1 - u_0, (i, j) \in \mathcal{M}; \end{aligned}$$

ceci montre que H_{1ij} est Lipschitz-continue avec une constante de Lipschitz de p_{ij}^{-1} . De façon similaire, H_{2ij} est Lipschitz-continue avec une constante de Lipschitz de p_{ij}^{-1} . Le théorème de Sklar implique que

$$P_C((0, u_1] \times (0, v] | A_{1ij} \times A_{2ij}) = C_{ij}(H_{1ij}(u), H_{2ij}(v))$$

avec

$$C_{ij}(u, v) = P_C((-\infty, H_{1ij}^{(-1)}(u)] \times (-\infty, H_{2ij}^{(-1)}(v)] | A_{1i} \times A_{2j}), (i, j) \in \mathcal{M}.$$

L'équation (3.2.2) et l'équation (3.2.3) sont automatiquement satisfaites.

(partie c, condition nécessaire) Si l'équation (3.2.2) et l'équation (3.2.3) sont satisfaites alors $H_{kij} \in \mathcal{H}$, $k = 1, 2$, $(i, j) \in \mathcal{M}$. Soit $\Phi_{1ij} = H_{1ij}$, $(i, j) \in \mathcal{M}$. Dans ce cas, $\Phi_{1i\cdot} = H_{1i\cdot}$ pour $P_C(U \in A_{1i}) > 0$, $U \sim \mathcal{U}(0, 1)$, $i \in J_1$. Si $i \in J_1$ satisfait $P_C(U \in A_{1i}) > 0$, $U \sim \mathcal{U}(0, 1)$, alors $\Phi_{1i\cdot} \circ \Phi_{1i\cdot}^{(-1)}(x) = x$ pour tout

$x \in (0, 1)$ et

$$\sum_{\{j:(i,j) \in \mathcal{M}\}} p_{ij} [H_{1ij} - \Phi_{1ij} \circ \Phi_{1i.}^{(-1)} \circ \Phi_{1i.}] = 0. \quad (3.2.4)$$

De plus nous avons toujours $\Phi_{1i.}^{(-1)} \circ \Phi_{1ij}(x) \leq x$, $x \in \mathbb{R}$, $(i, j) \in \mathcal{M}$. En utilisant cette propriété, nous obtenons $H_{1ij} \geq H_{1ij} \circ \Phi_{1i.}^{(-1)} \circ \Phi_{1i.} = \Phi_{1ij} \circ \Phi_{1i.}^{(-1)} \circ \Phi_{1i.}$, $(i, j) \in \mathcal{M}$; ainsi tous les éléments de l'équation (3.2.4) sont positifs ou nuls. Par conséquent, $H_{1ij} = \Phi_{1ij} \circ \Phi_{1i.}^{(-1)} \circ \Phi_{1i.}$ pour tous $(i, j) \in \mathcal{M}$.

De façon similaire, on peut montrer que si $\Phi_{2ij} = H_{2ij}$, $(i, j) \in \mathcal{M}$ alors $H_{2ij} = \Phi_{2ij} \circ \Phi_{2.j}^{(-1)} \circ \Phi_{2.j}$ pour tous $(i, j) \in \mathcal{M}$.

(partie c, condition suffisante) Si $\Phi_{kij} \in \mathcal{H}$, $k = 1, 2$, $(i, j) \in \mathcal{M}$ alors $\Phi_{1i.} \in \mathcal{H}$ si $P(U \in A_{1i}) > 0$, $i \in J_1$. (cela peut être prouvé par le théorème de convergence dominée). De façon similaire, $\Phi_{2.j} \in \mathcal{H}$ si $P(U \in A_{2j}) > 0$, $j \in J_2$. Si $P(U \in A_{1i}) > 0$, $i \in J_1$, et $H_{1ij} = \Phi_{1ij} \circ \Phi_{1i.}^{(-1)} \circ H_{1i.}$ alors $\Phi_{1i.} \circ \Phi_{1i.}^{(-1)}(x) = x$ pour tout $x \in (0, 1)$ et

$$\frac{\sum_{\{k:(i,k) \in \mathcal{M}\}} p_{ik} H_{1ik}}{\sum_{\{k:(i,k) \in \mathcal{M}\}} p_{ik}} = \Phi_{1i.} \circ \Phi_{1i.}^{(-1)} \circ H_{1i.} = H_{1i.}.$$

Cela montre que l'équation (3.2.2) est satisfaite. De façon similaire, nous pouvons montrer que l'équation (3.2.3) est satisfaite. \square

Théorème 3.2.1. (1) Une copule C est un prolongement d'une sous-copule D si et seulement si C satisfait

- (a) les conditions spécifiées au lemme 3.2.1 (partie 1) avec $A_{1i} = E_{1i}$, $A_{2j} = E_{2j}$, $(i, j) \in J_1 \times J_2$;
- (b) $p_{ij} = P_{\overline{D}}(E_{1i} \times E_{2j})$, $(i, j) \in J_1 \times J_2$;
- (c)

$$H_{10j}(u) = P_{\overline{D}}([0, u] \times [0, 1] | E_{1i} \times E_{2j}), \quad (3.2.5)$$

$u \in [0, 1]$, pour $(0, j) \in \mathcal{M}$,

$$H_{2i0}(v) = P_{\overline{D}}([0, 1] \times [0, v] | E_{1i} \times E_{2j}), \quad (3.2.6)$$

$v \in [0, 1]$, pour $(i, 0) \in \mathcal{M}$.

(d)

$$C_{00}(u, v) = P_C((-\infty, H_{100}^{(-1)}(u)] \times (-\infty, H_{200}^{(-1)}(v)] | E_{10} \times E_{20}),$$

$(0, 0) \in \mathcal{M}$, $u, v \in [0, 1]$.

- (2) Soit $\mathcal{M}^* = \{(i, j) \in \mathcal{M} : i, j > 0\}$. Soit C un prolongement de D , P_C la mesure de probabilité associée à C . Considérons les fonctions de répartition H_{1i+} et H_{2+j} données par

$$H_{1i+}(u) = \frac{P(U \leq u, U \in E_{1i}) - p_{i0}H_{1i0}(u)}{\sum_{j:(i,j) \in \mathcal{M}_+} p_{ij}},$$

$$P(U \in E_{1i}) > p_{i0}, i \in J_1^*,$$

$$H_{2+j}(u) = \frac{P(U \leq v, U \in E_{2j}) - p_{i0}H_{20j}(v)}{\sum_{j:(i,j) \in \mathcal{M}_+} p_{ij}},$$

$P(U \in E_{2j}) > p_{0j}, j \in J_2^*$. Les équations (3.2.2), (3.2.3) (avec $A_{kij} = E_{kij}$, $k = 1, 2$, $(i, j) \in J_1 \times J_2$ sont satisfaites si et seulement si les équations (3.2.5) et (3.2.6) sont satisfaites,

$$H_{1ij} = \Phi_{1ij} \circ \Phi_{1i+}^{(-1)} \circ H_{1i+}$$

avec

$$\phi_{1i+} = \frac{\sum_{\{k:(i,k) \in \mathcal{M}_+\} p_{ik} \phi_{1ik}}{\sum_{\{k:(i,k) \in \mathcal{M}_+\} p_{ik}},$$

$$(i, j) \in \mathcal{M}^*,$$

$$H_{2ij} = \Phi_{2ij} \circ \Phi_{2+j}^{(-1)} \circ H_{2+j},$$

avec

$$\phi_{2+j} = \frac{\sum_{\{k:(k,j) \in \mathcal{M}_+\} p_{kj} \phi_{2kj}}{\sum_{\{k:(k,j) \in \mathcal{M}_+\} p_{kj}},$$

$$(i, j) \in \mathcal{M}^*, \text{ pour } \phi_{1ik}, \phi_{2ij} \in \mathcal{H}, (i, j) \in \mathcal{M}^*.$$

DÉMONSTRATION. Considérons les fonctions $l_k : [0, 1] \rightarrow \mathbb{R}$ données par $l_k(t) = \sup\{u \in \overline{S}_k : u \leq t\}$, $t \in [0, 1]$, $k = 1, 2$. Nous avons $l_k(t) = t$, $t \in [0, 1]$ si et seulement si $t \in \overline{S}_k$, $k = 1, 2$. Si $l_k(t) < t$, $t \in [0, 1]$, alors $(l_k(t), t) \cap E_{k0} = \emptyset$, $k = 1, 2$. Cela implique que $[0, u] \cap E_{10} \times [0, 1] = [0, l_1(u)] \cap E_{10} \times [0, 1] \in \sigma(\mathcal{A})$ pour tout $u \in [0, 1]$ et $[0, 1] \times [0, v] \cap E_{20} = [0, 1] \times [0, l_2(v)] \cap E_{20} \in \sigma(\mathcal{A})$ pour tout $v \in [0, 1]$. Ainsi donc, $H_{10j}(u)$ et $H_{2i0}(v)$, $u, v \in [0, 1]$, $(i, j) \in \mathcal{M}$ sont bien définis même si $u \notin \overline{S}_1$ ou $v \notin \overline{S}_2$. Si $C \in \mathcal{C}$ alors les conditions données par le lemme (3.2.1) (partie 1) sont des conditions nécessaires. Si $C \in \mathcal{C}(D)$, P_C est la mesure de probabilité associée avec C et $P_{\overline{D}}$ est la mesure associée à \overline{D} sur $\sigma(\mathcal{A})$, alors nous devons avoir $P_C = P_{\overline{D}}$ sur $\sigma(\mathcal{A})$. Aussi, les autres conditions énoncées dans le théorème 3.2.1 sont des conditions nécessaires.

Soit $\mathcal{M}^* = \{(i, j) \in \mathcal{M} : i, j > 0\}$. Si les conditions du lemme 3.2.3 (partie 1) sont satisfaites alors $C \in \mathcal{C}$. Nous avons besoin de vérifier que la restriction de C sur $\overline{S}_1 \times \overline{S}_2$ est donnée par D . Notons que $[0, u] \cap E_{1i} \in \{\emptyset, E_{1i}\}$ pour tout

$u \in \overline{S}_1$, $(i, 0) \in \mathcal{M}$, $i > 0$. De façon similaire, $[0, v] \cap E_{2j} \in \{\emptyset, E_{2j}\}$ pour tout $v \in \overline{S}_2$, $(0, j) \in \mathcal{M}$, $j > 0$. Cela montre que

$$P_C([0, u] \times [0, v] | E_{1i} \times E_{2j}) = P_{\overline{D}}([0, u] \times [0, v] | E_{1i} \times E_{2j}), \quad (3.2.7)$$

$(u, v) \in \overline{S}_1 \times \overline{S}_2$, $(i, j) \in \mathcal{M}^*$. Si $v \in \overline{S}_2$, $E_{2j} \cap [0, v] = \emptyset$, $(0, j) \in \mathcal{M}$, $j > 0$ alors

$$P_C([0, u] \times [0, v] | E_{10} \times E_{2j}) = 0 = P_{\overline{D}}([0, u] \times [0, v] | E_{10} \times E_{2j}),$$

$u \in \overline{S}_1$.

Si $v \in \overline{S}_2$, $E_{2j} \cap [0, v] = E_{2j}$, $(0, j) \in \mathcal{M}$, $j > 0$ alors

$$P_C([0, u] \times [0, v] | E_{10} \times E_{2j}) = H_{10j}(u) = P_{\overline{D}}([0, u] \times [0, v] | E_{10} \times E_{2j}),$$

$u \in \overline{S}_1$.

Nous obtenons que

$$P_C([0, u] \times [0, v] | E_{10} \times E_{2j}) = P_{\overline{D}}([0, u] \times [0, v] | E_{10} \times E_{2j}), \quad (3.2.8)$$

$(u, v) \in \overline{S}_1 \times \overline{S}_2$, $(0, j) \in \mathcal{M}$, $j > 0$.

De façon similaire,

$$P_C([0, u] \times [0, v] | E_{1i} \times E_{20}) = P_{\overline{D}}([0, u] \times [0, v] | E_{1i} \times E_{20}), \quad (3.2.9)$$

$(u, v) \in \overline{S}_1 \times \overline{S}_2$, $(i, 0) \in \mathcal{M}$, $i > 0$.

De la définition de H_{k00} , $k = 1, 2$ et C_{00} , nous obtenons automatiquement que

$$P_C([0, u] \times [0, v] | E_{10} \times E_{20}) = P_{\overline{D}}([0, u] \times [0, v] | E_{10} \times E_{20}), \quad (3.2.10)$$

$(u, v) \in \overline{S}_1 \times \overline{S}_2$, $(0, 0) \in \mathcal{M}$.

En utilisant l'équation (3.2.7), l'équation (3.2.8), l'équation (3.2.9), l'équation (3.2.10), nous obtenons que

$$\begin{aligned} C(u, v) &= \sum_{(i,j) \in \mathcal{M}} P_C([0, u] \times [0, v] | E_{1i} \times E_{2j}) P_C(E_{1i} \times E_{2j}) \\ &= \sum_{(i,j) \in \mathcal{M}} P_{\overline{D}}([0, u] \times [0, v] | E_{1i} \times E_{2j}) P_{\overline{D}}(E_{1i} \times E_{2j}) \\ &= D(u, v), (u, v) \in \overline{S}_1 \times \overline{S}_2. \end{aligned}$$

La démonstration de la partie 2 est vraiment similaire à celle de la partie du lemme 3.2.1 et est omise. \square

3.3. LA COPULE MAXIMALE

Soit D une sous-copule. Dans cette section, nous trouvons C^{\max} , le prolongement maximal de D . Le résultat principal de cette partie est donné par le théorème 3.3.1. Le théorème 3.3.2 fait le lien entre le théorème 3.3.1 et la section 3.2. Il fait aussi le lien avec les autres résultats de la littérature scientifique.

Théorème 3.3.1. *Soit*

$$\begin{aligned} l_k(s) &= \sup\{t \in \overline{S}_k : t \leq s\}, s \in [0, 1], k = 1, 2 \\ L_k(s) &= \sup\{t \in \overline{S}_k : t \geq s\}, s \in [0, 1], k = 1, 2 \\ E_k(s) &= (l_k(s), L_k(s)], s \in [0, 1], k = 1, 2 \end{aligned}$$

$$C^{\max}(u, v) = \begin{cases} \overline{D}(u, v) \text{ si } (u, v) \in E_{10} \times E_{20} \\ \min(\overline{D}(b_{1i}, v), u - a_{1i} + \overline{D}(a_{1i}, v)) \text{ si } (u, v) \in E_{1i} \times E_{20} \\ \min(\overline{D}(u, b_{2j}), v - a_{2j} + \overline{D}(u, a_{2j})) \text{ si } (u, v) \in E_{10} \times E_{2j} \\ \min(C^{\max}(u, a_{2j}) + C^{\max}(a_{1i}), v) \\ - \overline{D}(a_{1i}, b_{2j}), C^{\max}(u, b_{2j}) + C^{\max}(a_{1i}), v) \\ - \overline{D}(a_{1i}, b_{2j}) \text{ si } (u, v) \in E_{1i} \times E_{2j}. \end{cases}$$

Nous avons

$$\begin{aligned} C^{\max}(u, v) &= \overline{D}(l_1(u), l_2(v)) + \min(P(U \leq u - l_1(u)), P_{\overline{D}}(E_1(u) \times [0, l_2(v)])) \\ &+ \min(\min(P(U \leq v - l_2(v)), P_{\overline{D}}([0, L_1(u)] \times E_2(v))), \\ &\quad \min(P(U \leq u - l_1(u)), P_{\overline{D}}(E_1(u) \times [0, L_1(v)])) \\ &+ \min(P(U \leq v - l_2(v)), P_{\overline{D}}([0, l_1(u)] \times E_2(v))). \end{aligned}$$

Nous avons les résultats suivants :

- (1) C^{\max} est une copule prolongeant D ,
- (2) $C \leq C^{\max}$, pour tout $C \in \mathcal{C}(D)$,
- (3) $C \in \mathcal{C}(D)$.

DÉMONSTRATION. Nous savons déjà que la restriction de C^{\max} sur $E_{10} \times E_{20}$ est donnée par la restriction de \overline{D} sur $E_{10} \times E_{20}$. Il est facile de vérifier que

$$C^{\max}(u, v) = \overline{D}(u, v) \text{ si } \begin{cases} u = b_{1i}, v \in E_{20}, i \in J_1^* \\ u \in E_{10}, v = b_{2j}, j \in J_2^* \\ (u, v) = (b_{1i}, b_{2j}), (i, j) \in J_1^* \times J_2^*. \end{cases}$$

Ceci montre que la restriction de C^{\max} sur $\overline{S}_1 \times \overline{S}_2$ est donnée par \overline{D} . Soit $C \in \mathcal{C}(D)$. Soit P_C la mesure de probabilité associée à C . Si $(u, v) \in E_{1i} \times E_{20}$, $i \in J_1^*$ alors

$$C(u, v) - \overline{D}(a_{1i}, v) = P_C((a_{1i}, u] \times [0, v]) \leq P_C(E_{1i} \times [0, v]),$$

$$C(u, v) - \overline{D}(a_{1i}, v) = P_C((a_{1i}, u] \times [0, v]) \leq P_C((a_{1i}, u] \times [0, 1]),$$

et $C(u, v) \leq C^{\max}(u, v)$. De la même façon, si $(u, v) \in E_{10} \times E_{2j}$, $j \in J_2^*$ alors $C(u, v) \leq C^{\max}(u, v)$. Si $(u, v) \in E_{1i} \times E_{2j}$, $(i, j) \in J_1^* \times J_2^*$ alors

$$\begin{aligned} C(u, v) &= P_C([0, u] \times [0, v]), \\ &\leq P_C([0, u] \times [0, a_{2j}]) + P_C([0, b_{1i}] \times (a_{2j}, v]), \\ &= C(u, a_{2j}) + (C(b_{1i}, v) - C(b_{1i}, a_{2j})), \\ &\leq C^{\max}(u, a_{2j}) + (C^{\max}(b_{1i}, v) - \overline{D}(b_{1i}, a_{2j})). \end{aligned}$$

Par symétrie, nous avons aussi

$$C(u, v) \leq C^{\max}(a_{1i}, v) + (C^{\max}(u, b_{2j}) - \overline{D}(a_{1i}, b_{2j})).$$

Ceci montre que $C \leq C^{\max}$. Montrons que $C \in \mathcal{C}$. Nous avons

$$\begin{aligned} C^{\max}(u, v) &= 0 \text{ si } \min(u, v) = 0, \\ C^{\max}(u, 1) &= u \text{ si } u \in [0, 1], \\ C^{\max}(1, v) &= v \text{ si } v \in [0, 1]. \end{aligned}$$

Il reste à prouver que la fonction volume $V_{C^{\max}}$ est positive. La fonction $V_{C^{\max}}$ est finiment additive. Tout rectangle $(u, u'] \times (v, v'] \subset [0, 1]^2$ peut être décomposée en une partition finie utilisant des rectangles de la forme $(u_1, u_2] \times (v_1, v_2]$ avec

$$(u_1, u_2] \subset E_{1i}, (v_1, v_2] \subset E_{2j}, (i, j) \in J_1^* \times J_2^* \quad (3.3.1)$$

$$u_1, u_2 \in \overline{S}_1, (v_1, v_2] \subset E_{2j}, j \in J_2^* \quad (3.3.2)$$

$$(u_1, u_2] \subset E_{1i}, v_1, v_2 \in \overline{S}_2, i \in J_1^* \quad (3.3.3)$$

$$u_1, u_2 \in \overline{S}_1, v_1, v_2 \in \overline{S}_2 \quad (3.3.4)$$

Pour les cas (3.3.1) à (3.3.4), nous avons besoin de montrer que $V_{C^{\max}}((u_1, u_2] \times (v_1, v_2]) \geq 0$. Soit

$$G^{\max}(i, j; u, v) = \min(V_{C^{\max}}((a_{1i}, u] \times E_{2j}), V_{C^{\max}}(E_{1i} \times (a_{2j}, v]))$$

pour $(u, v) \in \overline{E}_{1i} \times E_{2j}$, $(i, j) \in J_1^* \times J_2^*$. Nous avons

$$\begin{aligned} V_{C^{\max}}((u_1, u_2] \times (v_1, v_2]) &= G^{\max}(i, j; u_2, v_2) - G^{\max}(i, j; u_1, v_2) - G^{\max}(i, j; u_2, v_1) \\ &\quad + G^{\max}(i, j; u_1, v_1), \end{aligned}$$

pour $u_1, u_2 \in \overline{E}_{1i}, v_1, v_2 \in \overline{E}_{2j}, (i, j) \in J_1^* \times J_2^*$. La fonction $V_{C^{\max}}((a_{1i}, \cdot] \times E_{2j})$ est croissante sur $\overline{E}_{1i}, i \in J_1^*$, et la fonction $V_{C^{\max}}(E_{1i} \times (a_{2j}, \cdot])$ est croissante sur $\overline{E}_{2j}, j \in J_2^*$. Ceci implique que la fonction $G^{\max}(i, j; u_2, \cdot) - G^{\max}(i, j; u_1, \cdot)$ est croissante sur $\overline{E}_{2j}, j \in J_2^*$ pour tout $u_1 < u_2, u_1, u_2 \in \overline{E}_{1i}$ montrant par là que $V_{C^{\max}}$ est positive pour le cas (3.3.1). Nous avons aussi

$$C^{\max}(u, v) = \min(\overline{D}(u, b_{2j}), v - a_{2j} + \overline{D}(u, a_{2j})), (u, v) \in \overline{S}_1 \times \overline{E}_{2j}, j \in J_2^*.$$

La fonction $C^{\max}(u_2, \cdot) - C^{\max}(u_1, \cdot)$ est croissante sur \overline{E}_{2j} pour tout $u_1 < u_2, u_1, u_2 \in \overline{S}_1$, montrant ainsi que $V_{C^{\max}}$ est positive pour le cas (3.3.2). De la même façon, nous pouvons montrer que $V_{C^{\max}}$ est positive pour le cas (3.3.3). Enfin, $V_{C^{\max}}$ est positive pour le cas (3.3.4) car C^{\max} est un prolongement de \overline{D} . \square

Théorème 3.3.2. Posons

$$\begin{aligned} E_{1ij}^{\max} &= (a_{1i} + \overline{D}(a_{1i}, b_{2j}) - \overline{D}(a_{1i}, a_{2j}), a_{1i} + \overline{D}(b_{1i}, b_{2j}) - \overline{D}(b_{1i}, a_{2j})), \\ E_{1i0}^{\max} &= E_{1i} \setminus \bigcup_{j \in J_2^*} E_{1ij}^{\max}, \\ E_{2ij}^{\max} &= (a_{2j} + \overline{D}(b_{1i}, a_{2j}) - \overline{D}(a_{1i}, a_{2j}), a_{2j} + \overline{D}(b_{1i}, b_{2j}) - \overline{D}(a_{1i}, b_{2j})), \\ E_{20j}^{\max} &= E_{2j} \setminus \bigcup_{i \in J_1^*} E_{2ij}^{\max}, \end{aligned}$$

$i \in J_1^*, j \in J_2^*$.

Si C est le prolongement maximal, alors

$$\begin{aligned} C_{ij}(u, v) &= \min(u, v), u, v \in [0, 1], i + j > 0, (i, j) \in \mathcal{M}, \\ H_{1ij}(u) &= P(U \leq u | U \in E_{1ij}^{\max}), u \in [0, 1], i > 0, (i, j) \in \mathcal{M}, \\ H_{2ij}(v) &= P(U \leq v | U \in E_{2ij}^{\max}), v \in [0, 1], j > 0, (i, j) \in \mathcal{M} \end{aligned}$$

DÉMONSTRATION. Soit C le prolongement maximal. Si $(i, j) \in \mathcal{M}^*, (u, v) \in E_{1i} \times E_{2j}$, alors

$$\begin{aligned} p_{ij}H_{ij}(u, v) &= C(u, v) - C(u, a_{2j}) - C(a_{1i}, v) + C(a_{1i}, a_{2j}), \\ &= \min(C(b_{1i}, v) - C(b_{1i}, a_{2j}) - C(a_{1i}, v) + C(a_{1i}, a_{2j}), \\ &\quad C(u, b_{2j}) - C(u, a_{2j}) - C(a_{1i}, b_{2j}) + C(a_{1i}, a_{2j})), \\ &= p_{ij} \min(H_{1ij}(u), H_{2ij}(v)). \end{aligned}$$

Si $(0, j) \in \mathcal{M}, j > 0, (u, v) \in E_{1i} \times E_{2j}$, alors

$$\begin{aligned} p_{0j}H_{0j}(u, v) &= C(u, v) - C(u, a_{2j}) - \sum_{\{i>0:(i,j) \in \mathcal{M}\}} p_{ij}H_{ij}(u, v) \\ &= \min(\overline{D}(u, b_{2j}) - \overline{D}(u, a_{2j}), v - a_{2j}) - \sum_{\{i>0:b_{1i} \leq u, (i,j) \in \mathcal{M}\}} p_{ij}H_{2ij}(v), \end{aligned}$$

avec

$$\begin{aligned} p_{0j}H_{10j}(u) &= \overline{D}(u, b_{2j}) - \overline{D}(u, a_{2j}) - \sum_{\{i>0: b_{1i} \leq u, (i,j) \in \mathcal{M}\}} p_{ij}, \\ p_{0j}H_{20j}(v) &= v - a_{2j} - \sum_{\{i>0: (i,j) \in \mathcal{M}\}} p_{ij}H_{2ij}(v). \end{aligned}$$

Soit $c_j^{\max}(u) = a_{2j} + \overline{D}(u, b_{2j}) - \overline{D}(u, a_{2j})$, $u \in [0, 1]$, $(0, j) \in \mathcal{M}$. Nous obtenons

$$H_{2ij}(c_j^{\max}(u)) = I_{[b_{1i}, \infty)}(u), u \in E_{10}, (i, j) \in \mathcal{M}^*.$$

Si $v \geq c_j^{\max}(u)$, $(u, v) \in E_{10} \times E_{2j}$, $(0, j) \in \mathcal{M}$, $j > 0$ alors $H_{2ij}(v) = 1$ pour $b_{1j} \leq u$ et $H_{0j}(u, v) = H_{10j}(u) = H_{20j}(c_j^{\max}(u)) \leq H_{20j}(v)$, donc $H_{0j}(u, v) = \min(H_{10j}(u), H_{20j}(v))$.

Si $v < c_j^{\max}(u)$, $(u, v) \in E_{10} \times E_{2j}$, $(0, j) \in \mathcal{M}$, $j > 0$ alors $H_{2ij}(v) = 0$ pour $b_{1j} < u$ et $H_{0j}(u, v) = H_{20j}(v) \leq H_{20j}(c_j^{\max}(u)) \leq H_{10j}(u)$, donc $H_{0j}(u, v) = \min(H_{10j}(u), H_{20j}(v))$. De la même façon, nous pouvons montrer que

$$H_{i0}(u, v) = \min(H_{1i0}(u), H_{2i0}(v)), i > 0, (i, 0) \in \mathcal{M}.$$

□

3.4. LA COPULE MINIMALE

Soit D une sous-copule. Dans cette section, nous trouvons C^{\min} , le prolongement minimal de D . Le résultat principal de cette partie est donné par le théorème 3.4.1. Le théorème 3.4.2 fait le lien entre le théorème 3.4.1 et la section 3.2. Il fait aussi le lien avec les autres résultats de la littérature scientifique.

Théorème 3.4.1. *Soit*

$$\begin{aligned} l_k(s) &= \sup\{t \in \overline{S}_k : t \leq s\}, s \in [0, 1], k = 1, 2 \\ L_k(s) &= \sup\{t \in \overline{S}_k : t \geq s\}, s \in [0, 1], k = 1, 2 \\ E_k(s) &= (l_k(s), L_k(s)], s \in [0, 1], k = 1, 2 \end{aligned}$$

$$C^{\min}(u, v) = \begin{cases} \overline{D}(u, v) si (u, v) \in E_{10} \times E_{20} \\ \max(\overline{D}(a_{1i}, v), u - b_{1i} + \overline{D}(b_{1i}, v)) si (u, v) \in E_{1i} \times E_{20} \\ \max(\overline{D}(u, a_{2j}), v - b_{2j} + \overline{D}(u, b_{2j})) si (u, v) \in E_{10} \times E_{2j} \\ \max(C^{\min}(u, a_{2j}) + C^{\min}(a_{1i}), v) \\ -\overline{D}(a_{1i}, a_{2j}), C^{\min}(u, b_{2j}) + C^{\min}(b_{1i}), v) \\ -\overline{D}(b_{1i}, b_{2j})) si (u, v) \in E_{1i} \times E_{2j} \end{cases}$$

Nous avons les résultats suivants :

- (1) C^{\min} est une copule prolongeant D ,
- (2) $C \geq C^{\min}$, pour tout $C \in \mathcal{C}(D)$,
- (3) $C^{\min} \in \mathcal{C}(D)$.

DÉMONSTRATION. Nous savons déjà que la restriction de C^{\min} sur $E_{10} \times E_{20}$ est donnée par la restriction de \overline{D} sur $E_{10} \times E_{20}$. Il est facile de vérifier que

$$C^{\min}(u, v) = \overline{D}(u, v) \text{ si } \begin{cases} u = b_{1i}, v \in E_{20}, i \in J_1^* \\ u \in E_{10}, v = b_{2j}, j \in J_2^* \\ (u, v) = (b_{1i}, b_{2j}), (i, j) \in J_1^* \times J_2^*. \end{cases}$$

Ceci montre que la restriction de C^{\max} sur $\overline{S}_1 \times \overline{S}_2$ est donnée par \overline{D} . Soit $C \in \mathcal{C}(D)$. Soit P_C la mesure de probabilité associée à C . Si $(u, v) \in E_{1i} \times E_{20}$, $i \in J_1^*$ alors

$$\overline{D}(b_{1i}, v) - C(u, v) = P_C((u, b_{1i}] \times [0, v]) \leq \overline{D}(b_{1i}, v) - \overline{D}(a_{1i}, v),$$

$$\overline{D}(b_{1i}, v) - C(u, v) = P_C((u, b_{1i}] \times [0, 1]) \leq b_{1i} - u,$$

et $C(u, v) \geq C^{\min}(u, v)$. De la même façon, si $(u, v) \in E_{10} \times E_{2j}$, $j \in J_2^*$ alors $C(u, v) \geq C^{\min}(u, v)$. Si $(u, v) \in E_{1i} \times E_{2j}$, $(i, j) \in J_1^* \times J_2^*$ alors

$$\begin{aligned} C(u, v) &= P_C([0, u] \times [0, v]), \\ &\leq P_C([0, a_{1j}] \times [0, v]) + P_C([0, u] \times (0, a_{2j}]) - P_C([0, a_{1j}] \times (0, a_{2j}]), \\ &= C(u, a_{2j}) + (C(a_{1i}, v) - C(a_{1i}, a_{2j})), \\ &\geq C^{\min}(u, a_{2j}) + (C^{\min}(a_{1i}, v) - \overline{D}(a_{1i}, a_{2j})). \end{aligned}$$

Par symétrie, nous avons aussi

$$C(u, v) \leq C^{\min}(b_{1i}, v) + (C^{\min}(u, b_{2j}) - \overline{D}(b_{1i}, b_{2j})).$$

Ceci montre que $C \geq C^{\min}$. Montrons que $C^{\min} \in \mathcal{C}$. Nous avons

$$\begin{aligned} C^{\min}(u, v) &= 0 \text{ si } \min(u, v) = 0, \\ C^{\min}(u, 1) &= u \text{ si } u \in [0, 1], \\ C^{\min}(1, v) &= v \text{ si } v \in [0, 1]. \end{aligned}$$

Il reste à prouver que la fonction volume $V_{C^{\min}}$ est positive. La fonction $V_{C^{\min}}$ est finiment additive. Tout rectangle $(u, u'] \times (v, v'] \subset [0, 1]^2$ peut être décomposée en une partition finie utilisant des rectangles de la forme $(u_1, u_2] \times (v_1, v_2]$ avec

$$(u_1, u_2] \subset E_{1i}, (v_1, v_2] \subset E_{2j}, (i, j) \in J_1^* \times J_2^* \quad (3.4.1)$$

$$u_1, u_2 \in \overline{S}_1, (v_1, v_2] \subset E_{2j}, j \in J_2^* \quad (3.4.2)$$

$$(u_1, u_2] \subset E_{1i}, v_1, v_2 \in \overline{S}_2, i \in J_1^* \quad (3.4.3)$$

$$u_1, u_2 \in \overline{S}_1, v_1, v_2 \in \overline{S}_2 \quad (3.4.4)$$

Pour les cas (3.4.1) à (3.4.4), nous avons besoin de montrer que $V_{C^{\min}}((u_1, u_2] \times (v_1, v_2]) \geq 0$. Soit

$$G^{\min}(i, j; u, v) = -\min(V_{C^{\min}}((u, b_{1i}] \times E_{2j}), V_{C^{\min}}(E_{1i} \times (a_{2j}, v]))$$

pour $(u, v) \in \overline{E}_{1i} \times E_{2j}$, $(i, j) \in J_1^* \times J_2^*$. Nous avons

$$V_{C^{\min}}((u_1, u_2] \times (v_1, v_2]) = G^{\min}(i, j; u_2, v_2) - G^{\min}(i, j; u_1, v_2) - G^{\min}(i, j; u_2, v_1) + G^{\max}(i, j; u_1, v_1),$$

pour $u_1, u_2 \in \overline{E}_{1i}$, $v_1, v_2 \in \overline{E}_{2j}$, $(i, j) \in J_1^* \times J_2^*$. La fonction $V_{C^{\min}}((\cdot, b_{1i}] \times E_{2j})$ est décroissante sur \overline{E}_{1i} , $i \in J_1^*$, et la fonction $V_{C^{\min}}(E_{1i} \times (a_{2j}, \cdot])$ est croissante sur \overline{E}_{2j} , $j \in J_2^*$. Ceci implique que la fonction $G^{\min}(i, j; u_2, \cdot) - G^{\max}(i, j; u_1, \cdot)$ est croissante sur \overline{E}_{2j} , $j \in J_2^*$ pour tout $u_1 < u_2$, $u_1, u_2 \in \overline{E}_{1i}$ montrant par là que $V_{C^{\min}}$ est positive pour le cas (3.4.1). Nous avons aussi

$$C^{\min}(u, v) = \min(\overline{D}(u, a_{2j}), v - b_{2j} + \overline{D}(u, b_{2j})), (u, v) \in \overline{S}_1 \times \overline{E}_{2j}, j \in J_2^*.$$

La fonction $C^{\min}(u_2, \cdot) - C^{\min}(u_1, \cdot)$ est croissante sur \overline{E}_{2j} pour tout $u_1 < u_2$, $u_1, u_2 \in \overline{S}_1$, montrant ainsi que $V_{C^{\min}}$ est positive pour le cas (3.4.2). De la même façon, nous pouvons montrer que $V_{C^{\min}}$ est positive pour le cas (3.4.3). Enfin, $V_{C^{\min}}$ est positive pour le cas (3.4.4) car C^{\min} est un prolongement de \overline{D} . \square

Théorème 3.4.2. Posons

$$\begin{aligned} E_{1ij}^{\min} &= (b_{1i} - \overline{D}(b_{1i}, b_{2j}) + \overline{D}(a_{1i}, b_{2j}), b_{1i} + \overline{D}(a_{1i}, a_{2j}) - \overline{D}(b_{1i}, a_{2j}]), \\ E_{1i0}^{\min} &= E_{1i} \setminus \bigcup_{j \in J_2^*} E_{1ij}^{\max}, \\ E_{2ij}^{\min} &= (b_{2j} + \overline{D}(b_{1i}, a_{2j}) - \overline{D}(b_{1i}, b_{2j}), b_{2j} + \overline{D}(a_{1i}, a_{2j}) - \overline{D}(a_{1i}, b_{2j}]), \\ E_{20j}^{\min} &= E_{2j} \setminus \bigcup_{i \in J_1^*} E_{2ij}^{\max}, \end{aligned}$$

$i \in J_1^*$, $j \in J_2^*$.

Si C est le prolongement minimal, alors

$$\begin{aligned} C_{ij}(u, v) &= \max(0, u + v - 1), u, v \in [0, 1], i + j > 0, (i, j) \in \mathcal{M}, \\ H_{1ij}(u) &= P(U \leq u | U \in E_{1ij}^{\min}), u \in [0, 1], i > 0, (i, j) \in \mathcal{M}, \\ H_{2ij}(v) &= P(U \leq v | U \in E_{2ij}^{\min}), v \in [0, 1], j > 0, (i, j) \in \mathcal{M} \end{aligned}$$

DÉMONSTRATION. Soit C le prolongement minimal. Si $(i, j) \in \mathcal{M}^*$, $(u, v) \in E_{1i} \times E_{2j}$, alors

$$\begin{aligned}
p_{ij}H_{ij}(u, v) &= C(u, v) - C(u, a_{2j}) - C(a_{1i}, v) + C(a_{1i}, a_{2j}), \\
&= \max(0, (C(b_{1i}, v) - C(a_{1i}, v)) + (C(u, b_{2j}) - C(u, a_{2j})) - \\
&\quad (C(b_{1i}, b_{2j}) - C(a_{1i}, a_{2j}))) \\
&= p_{ij} \max(0, H_{1ij}(u) + H_{2ij}(v) - 1).
\end{aligned}$$

Si $(0, j) \in \mathcal{M}$, $j > 0$, $(u, v) \in E_{10} \times E_{2j}$, alors

$$\begin{aligned}
p_{0j}H_{0j}(u, v) &= C(u, v) - C(u, a_{2j}) - \sum_{\{i>0:(i,j) \in \mathcal{M}\}} p_{ij}H_{ij}(u, v) \\
&= \max(0, v - b_{2j} + \bar{D}(u, b_{2j}) - \bar{D}(u, a_{2j}) - \sum_{\{i>0:b_{1i} \leq u, (i,j) \in \mathcal{M}\}} p_{ij}H_{2ij}(v)),
\end{aligned}$$

avec

$$\begin{aligned}
p_{0j}H_{10j}(u) &= \bar{D}(u, b_{2j}) - \bar{D}(u, a_{2j}) - \sum_{\{i>0:b_{1i} \leq u, (i,j) \in \mathcal{M}\}} p_{ij}, \\
p_{0j}H_{20j}(v) &= v - a_{2j} - \sum_{\{i>0:(i,j) \in \mathcal{M}\}} p_{ij}H_{2ij}(v).
\end{aligned}$$

Soit $c_j^{\min}(u) = b_{2j} - (\bar{D}(u, b_{2j}) - \bar{D}(u, a_{2j}))$, $u \in [0, 1]$, $(0, j) \in \mathcal{M}$. Nous obtenons

$$H_{2ij}(c_j^{\min}(u)) = I_{[-\infty, b_{1i}]}(u), u \in E_{10}, (i, j) \in \mathcal{M}^*.$$

Si $v \geq c_j^{\min}(u)$, $(u, v) \in E_{10} \times E_{2j}$, $(0, j) \in \mathcal{M}$, $j > 0$ alors $H_{2ij}(v) = 1$ pour $b_{1j} > u$ et $H_{0j}(u, v) = H_{10j}(u) + H_{20j}(v) - 1 \geq H_{10j}(u) + H_{20j}(c_j^{\min}(u)) - 1 = 0$, donc $H_{0j}(u, v) = \max(0, H_{10j}(u) + H_{20j}(v) - 1)$.

Si $v < c_j^{\min}(u)$, $(u, v) \in E_{10} \times E_{2j}$, $(0, j) \in \mathcal{M}$, $j > 0$ alors $H_{2ij}(v) = 0$ pour $b_{1j} \leq u$ et $H_{0j}(u, v) = 0 = H_{10j}(u) + H_{20j}(c_j^{\min}(u)) - 1 \geq H_{10j}(u) + H_{20j}(v) - 1$, donc $H_{0j}(u, v) = \min(H_{10j}(u), H_{20j}(v))$. De la même façon, nous pouvons montrer que

$$H_{i0}(u, v) = \max(0, H_{1i0}(u) + H_{2i0}(v) - 1), i > 0, (i, 0) \in \mathcal{M}.$$

□

3.5. LA MÉTHODE DES RECTANGLES ET L'ENSEMBLE DE CANTOR

Cette section est motivée par l'article de [de Amo et coll., 2012]. Les ensembles P_1 et P_2 apparaissant dans [de Amo et coll., 2012] sont liés aux ensembles N_1 et N_2 définis dans cette section. L'approche développée dans [de Amo et coll., 2012] est basée sur la méthode des rectangles. Dans la méthode des rectangles, D est une sous-copule sur $S_1 \times S_2$. La partition de $[0, 1]$ a la forme :

$$[0, 1] = N_1 \cup \left\{ \bigcup_{i \in J_1} E_{1i} \right\}, J_1 \subset \mathbb{N},$$

les ensembles E_{1i} , $i \in J_1$ sont des intervalles et l'ensemble N_1 est appelé ensemble négligeable. Les ensembles E_{1i} , $i \in J_1$, satisfont les conditions $\partial E_{1i} \subset \bar{S}_1$ et $\dot{E}_{2j} \cap \bar{S}_2 \in \{\emptyset, \dot{E}_{2j}\}$.

Exemple 3.5.1. Si $S_1 = \{s_j : j = 1, \dots, n\}$ avec $0 = s_1 < \dots < s_n = 1$ alors nous pouvons prendre

$$N_1 = \{0\}, E_{1i} = (s_i, s_{i+1}], i = 1, \dots, n-1.$$

Exemple 3.5.2. Si X est une variable aléatoire discrète à valeurs dans \mathbb{N} ayant une fonction de répartition F avec $f(n) < 1$ pour tout $n \in \mathbb{N}$ et R est l'espace image de F alors nous pouvons prendre $S_1 = R \cup \{1\} = \{s_j : j = 1, 2, \dots\} \cup \{1\}$ avec $0 = s_1 < s_2 < \dots$,

$$N_1 = \{0, 1\}, E_{1i} = (s_i, s_{i+1}], i = 1, 2, \dots$$

En général, nous pouvons écrire

$$\begin{aligned} [0, 1] &= \partial S_1 \cup \{\cup_{i \in J_1} W_{1i}\}, J_1 \subset \mathbb{N} \\ &= N_1 \cup \{\cup_{i \in J_1} E_{1i}\}, \end{aligned}$$

avec $W_{1i} = (\alpha_{1i}, \beta_{1i})$ des intervalles disjoints et

$$E_{1i} = (\alpha_{1i}, \beta_{1i}], i \in J_1, N_1 = [0, 1] \setminus \{\cup_{i \in J_1} E_{1i}\}.$$

Dans la méthode des rectangles, nous voulons éliminer les ensembles négligeables. Pour ce faire, nous avons besoin de montrer que si C est un prolongement de D et P_C une mesure de probabilité associée à C , alors $P_C(\cup_{i \in J_1} \cup_{j \in J_2} E_{1i} \times E_{2j}) = 1$.

Lemme 3.5.1. Nous avons $P_C(\cup_{i \in J_1} \cup_{j \in J_2} E_{1i} \times E_{2j}) = 1$ si et seulement si $P(U \in \partial S_k) = 0, k = 1, 2$.

DÉMONSTRATION. Prouver que $P_C(\cup_{i \in J_1} \cup_{j \in J_2} E_{1i} \times E_{2j}) = 1$ est équivalent à prouver que $P_C(\{N_1 \times [0, 1]\} \cup \{[0, 1] \times N_2\}) = 0$. Nous avons

$$\begin{aligned} P_C(\{N_1 \times [0, 1]\} \cup \{[0, 1] \times N_2\}) &\leq P_C(\{N_1 \times [0, 1]\}) + P_C(\{[0, 1] \times N_2\}) \\ &= \sum_{k=1}^2 P(U \in \partial S_k) \end{aligned}$$

et

$$\begin{aligned} P_C(\{N_1 \times [0, 1]\} \cup \{[0, 1] \times N_2\}) &\geq \max(P_C(\{N_1 \times [0, 1]\}), P_C(\{[0, 1] \times N_2\})) \\ &= \max_{k \in \{1, 2\}} P(U \in \partial S_k). \end{aligned}$$

□

Si la sous-copule D est induite par un vecteur aléatoire discret (X_1, X_2) alors S_1 est l'espace image de la fonction de répartition de X_1 et $P(U \in \partial S_1) = 0$. Cependant, il est bien connu qu'il existe des ensembles $S \subset [0, 1]$, satisfaisant $P(U \in \partial S) > 0$. Nous pouvons prendre en exemple les ensembles de Cantor. Considérons $(\delta, \rho) \in (0, 1] \times (0, 1)$ et soit $F_\infty(\delta, \rho)$ l'ensemble de Cantor associé à (δ, ρ) . Soit μ la mesure de Lebesgue sur $[0, 1]$. Nous construisons $F_\infty(\delta, \rho)$ à travers une formule récursive. À l'étape 1, nous retranchons de $[0, 1]$ l'intervalle $V_{1,1}(\delta, \rho)$ de longueur $\delta(1 - \rho)$ centré en $1/2$. Nous obtenons la partition

$$\begin{aligned} [0, 1] &= [0, (1 - \delta\rho)/2] \cup ((1 - \delta\rho)/2, (1 + \delta\rho)/2) \cup [(1 + \delta\rho)/2, 1], \\ &= F_{1,1}(\delta, \rho) \cup V_{1,1}(\delta, \rho) \cup F_{1,2}(\delta, \rho). \end{aligned}$$

À l'étape n , nous avons la partition

$$\begin{aligned} [0, 1] &= F_{n,1}(\delta, \rho) \cup V_{n,1}(\delta, \rho) \cup F_{n,2}(\delta, \rho) \cup V_{n,2}(\delta, \rho) \cup F_{n,3}(\delta, \rho) \\ &\quad \cup \dots \cup V_{n,2^n-1}(\delta, \rho) \cup F_{n,2^n}(\delta, \rho). \end{aligned}$$

Les ensembles $F_{n,k}$, $k = 1, \dots, 2^n$ sont fermés,

$$\begin{aligned} \mu(F_{n,k}(\delta, \rho)) &= [1 - \delta(1 - \rho^n)]/2^n, k = 1, \dots, 2^n, \\ \mu(V_{n,2k-1}(\delta, \rho)) &= \delta(1 - \rho)\rho^{n-1}/2^n, k = 1, \dots, 2^{n-1}. \end{aligned}$$

À l'étape $n + 1$, les ensembles $F_{n,k}(\delta, \rho)$, $k = 1, \dots, 2^n$, sont partitionnés. Nous obtenons

$$\begin{aligned} F_{n,k}(\delta, \rho) &= F_{n+1,2k-1}(\delta, \rho) \cup V_{n+1,2k-1}(\delta, \rho) \cup F_{n+1,2k}(\delta, \rho), \\ &\quad k = 1, \dots, 2^n \\ V_{n+1,2k}(\delta, \rho) &= V_{n,2k}(\delta, \rho), k = 1, \dots, 2^n - 1. \end{aligned}$$

Nous posons $F_n(\delta, \rho) = \bigcup_{k=1}^{2^n} F_{n,k}(\delta, \rho)$. Les ensembles $F_n(\delta, \rho)$ sont fermés et emboîtés. Finalement, posons $F_\infty(\delta, \rho) = \bigcap_{n=1}^{\infty} F_n(\delta, \rho)$. L'ensemble $F_\infty(\delta, \rho)$ est un ensemble de Cantor. Cet ensemble est nul part dense, et $\mu(F_\infty(\delta, \rho)) = 1 - \delta > 0$ si $0 < \delta, \rho < 1$. Les cas où $\rho = 1/2$ correspondent aux ensembles classiques de Cantor dans la littérature.

Définition 3.5.1. Une variable aléatoire X sur $[0, 1]$ a la distribution de Cantor (δ, ρ) si X est de loi uniforme sur $[0, 1]$ pour $\delta = 0$, $\rho \in (0, 1)$ et

$$\begin{aligned} P(X \leq x) &= \sup\{y \in F_\infty(\delta, \rho) : y \leq x\}, x \in [0, 1], \\ &\quad (\delta, \rho) \in (0, 1] \times (0, 1) \end{aligned} \tag{3.5.1}$$

Nous pouvons voir que l'équation 3.5.1 donne une fonction de répartition car $F_\infty(\delta, \rho)$ est fermé. Soit $F_\infty(0, \rho) = [0, 1]$, $\rho \in (0, 1)$. Lorsque $X \sim \text{Cantor}(\delta, \rho)$, $0 \leq \delta \leq 1$, $\rho \in (0, 1)$, l'espace image de la fonction de répartition de X est donné par $F_\infty(\delta, \rho)$. Lorsque $\delta = 1$, $\rho \in (0, 1)$, X est une variable aléatoire discrète.

Considérons une sous-copule D associée au vecteur aléatoire (X_1, X_2) et soient F_1, F_2 les fonctions de répartition marginales associées. Soient α_{1i}, β_{1i} , $i \in J_1$, les bornes des ensembles E_{1i} , $i \in J_1$. Nous obtenons toujours que $P(F_1(X_1) = \beta_{1i}) > 0$, $i \in J_1$. Dans l'exemple 1, et dans l'exemple 2, il est aussi vrai que $P(F_1(X_1) = \alpha_{1i}) > 0$, $i \in J_1$. En général, nous pouvons obtenir que $P(F_1(X_1) = \alpha_{1i}) = 0$ pour un certain $i \in J_1$. Pour les cas discrets, nous pouvons penser que $\mu(\partial S_1) = 0$ parce que ∂S_1 est aussi au plus dénombrable mais ce n'est pas toujours le cas. Considérons $X \sim \text{Cantor}(1, 1/5)$. Nous avons

$$P(X = x) > 0 \quad \text{si et seulement si } x = 0, d_1 d_2 \cdots d_l \\ \text{avec } d_k \in \{0, 9\}, k = 1, \dots, l-1, d_l = 9, l \geq 1.$$

Si $u \in [0, 1] \setminus \overline{S}_1$, alors $u = 0, d_1 d_2 \cdots d_l$ avec $d_k \in \{0, 9\}$, $k = 1, \dots, l-1$, $d_l \in \{1, 2, \dots, 8\}$, $l \geq 1$,

$$l_1(u) = 0, d_1 d_2 \cdots d_{l-1} 1 \\ L_1(u) = 0, d_1 d_2 \cdots d_{l-1} 9$$

montrant par là que $P(F_1(X_1) = \alpha_{1i}) = 0$, $i \in J_1^*$. Finalement, un point $u \in \partial S_1$ si et seulement si

$$u = 0, d_1 d_2 d_3 \cdots \text{ avec } d_k \in \{0, 9\}, k \geq 1$$

montrant par là que ∂S_1 n'est pas dénombrable.

3.6. CONCLUSION

Pour une sous-copule donnée D , il s'est agi de caractériser dans le cadre le plus général possible, les prolongements de D . En dimension 2, il s'avère que les bornes supérieures et inférieures des prolongements sont aussi des prolongements de D . Ce résultat a été dans le passé atteint par [de Amo et coll., 2012]. Mais le résultat en question n'était pas le plus général possible ; la section 5 explique pourquoi.

Chapitre 4

BORNE SUPÉRIEURE DES PROLONGEMENTS D'UNE SOUS-COPULE À DOMAINE FINI EN DIMENSION 3

Résumé. *La recherche de la borne supérieure des prolongements d'une sous-copule discrète en dimension 3 est un problème posé par [Carley, 2002]. L'apport de ce chapitre est de donner en dimension 3 en tout point $(u_1, u_2, u_3) \in [0, 1]^3$ la formule de la borne supérieure.*

4.1. INTRODUCTION

Pour une sous-copule D , l'ensemble des prolongements possibles peut comporter un grand nombre d'éléments [Genest et Nešlehová, 2007]. Dans la pratique, le prolongement de [Schweizer et Sklar, 1974] ou ses variantes [Rüschendorf, 2009; Denuit et Lambert, 2005] sont les plus utilisées mais il se pose le problème d'un critère de choix d'un prolongement. À défaut de proposer un critère de sélection, plusieurs auteurs se sont penchés sur la détermination des bornes de l'ensemble des prolongements d'une sous-copule. Le problème se formule comme suit : étant donnée D une sous-copule, on étudie les propriétés des fonctions $UC : [0, 1]^d \rightarrow [0, 1]$ et $LC : [0, 1]^d \rightarrow [0, 1]$ données par

$$UC(u_1, \dots, u_d) = \sup\{C(u_1, \dots, u_d) : C \text{ est un prolongement de } D\},$$

$$LC(u_1, \dots, u_d) = \inf\{C(u_1, \dots, u_d) : C \text{ est un prolongement de } D\},$$

$u_1, \dots, u_d \in [0, 1], d \geq 2$.

Ce problème a été abordé pour la première fois par [Carley, 2002] dans le cas d'une sous-copule D avec un domaine de cardinalité finie. [Carley, 2002] a prouvé qu'en dimension 2, les fonctions UC et LC étaient des copules et a donné leurs expressions. Soient $R_1 = \{a_{11}, a_{12}, \dots, a_{1n_1}\}$, $R_2 = \{a_{21}, a_{22}, \dots, a_{2n_2}\}$, $0 = a_{11} < a_{12} < \dots < a_{1n_1} = 1$, $0 = a_{21} < a_{22} < \dots < a_{2n_2} = 1$. La solution proposée par

[Carley, 2002] et étendu au cas de domaine infini par [Genest et Nešlehová, 2007] est :

$$\begin{aligned}
UC(u_1, u_2) &= \sum_{j_1=2}^{n_1} \sum_{j_2=2}^{n_2} \max(0, \min(u_1 - a_{1j_1-1} - \sum_{l=1}^{j_2-1} p_{j_1 l}, \\
&\quad u_2 - a_{2j_2-1} - \sum_{l=1}^{j_1-1} p_{l j_2}, p_{j_1 j_2})) \\
LC(u_1, u_2) &= \sum_{j_1=2}^{n_1} \sum_{j_2=2}^{n_2} \max(0, \max(0, \min(u_1 - a_{1j_1-1} - \sum_{l=j_2+1}^n p_{j_1 l}, p_{j_1 j_2})) + \\
&\quad \max(0, \min(u_2 - a_{2j_2-1} - \sum_{l=j_1+1}^{n_2} p_{l j_2}, p_{j_1 j_2})) - p_{j_1 j_2})
\end{aligned}$$

avec $p_{j_1 j_2} = D(a_{1j_1}, a_{2j_2}) - D(a_{1j_1-1}, a_{2j_2}) - D(a_{1j_1}, a_{2j_2-1}) + D(a_{1j_1-1}, a_{2j_2-1})$, $j_1 \in \{2, \dots, n_1\}$, $j_2 \in \{2, \dots, n_2\}$.

Notons que la formulation présentée ici est celle de [Genest et Nešlehová, 2007]. La technique de résolution de [Carley, 2002], encore appelée principe de la gratification immédiate, repose sur les égalités suivantes :

$$\begin{aligned}
UC(a_{1j_1-1}, u_2) &= D(a_{1j_1-1}, a_{2j_2-1}) + \min(u_2 - a_{2j_2-1}, \sum_{l=1}^{j_1-1} p_{l j_2}), \\
UC(u_1, a_{2j_2-1}) &= D(a_{1j_1-1}, a_{2j_2-1}) + \min(u_1 - a_{1j_1-1}, \sum_{l=1}^{j_2-1} p_{j_1 l}), \\
LC(a_{1j_1-1}, u_2) &= D(a_{1j_1-1}, a_{2j_2-1}) + \min(u_2 - a_{2j_2-1} - \sum_{l>j_1-1} p_{l j_2}, \sum_{l=1}^{j_1-1} p_{l j_2}), \\
LC(u_1, a_{2j_2-1}) &= D(a_{1j_1-1}, a_{2j_2-1}) + \min(u_1 - a_{1j_1-1} - \sum_{l>j_2-1} p_{j_1 l}, \sum_{l=1}^{j_2-1} p_{j_1 l}),
\end{aligned} \tag{4.1.1}$$

$u_1 \in (a_{1j_1-1}, a_{1j_1}]$, $u_2 \in (a_{2j_2-1}, a_{2j_2}]$, $j_1 \in \{2, \dots, n_1\}$, $j_2 \in \{2, \dots, n_2\}$. Ces relations proviennent du fait que les bornes inférieures et supérieures sont des copules, et que les copules ont des marges uniformes. En dimension 3, il y a cependant deux difficultés à surmonter pour généraliser ces équations :

- il n'est pas sûr que la borne supérieure en dimension 3 soit une copule ;
- toute généralisation se heurte au problème d'incompatibilité ; en effet, il n'est pas toujours possible de construire une copule en dimension $d > 2$ avec des marges multivariées fixées.

L'étude de la borne supérieure des prolongements en dimension $d > 2$ nécessite l'introduction de la notion de quasi-copule. Les quasi-copules ont été introduites par [Alsina, Nelsen et Schweizer, 1993]. Une quasi-copule Q sur $[0, 1]^d$ est une fonction $Q : [0, 1]^d \rightarrow [0, 1]$ vérifiant selon [Molina et Sempi, 2005; Nelsen, 1999] les propriétés suivantes :

- Q est croissante par rapport à chaque variable ;
- $Q(u_1, \dots, u_d) = 0$ si $u_1 \cdots u_d = 0$, et $Q(u_1, \dots, u_d) = u_i$ si $u_1 \cdots u_d = u_i$, $u_i \in [0, 1]$, $i = 1, \dots, d$;

$$— |Q(u_1, \dots, u_d) - Q(u'_1, \dots, u'_d)| \leq \sum_{i=1}^d |u_i - u'_i|, u_i, u'_i \in [0, 1], i = 1, \dots, d.$$

Une autre caractérisation des quasi-copules est donnée par [Genest, Molina, Lallena et Sempi, 1999]. En dimension $d = 2$, [Nelsen et coll., 2008] ont montré que la borne supérieure des copules qui admettent la même restriction sur un sous-ensemble fixé de $[0, 1]^2$ est une quasi-copule. [Rodríguez-Lallena et Úbeda Flores, 2007] ont déterminé en dimension $d > 2$ la borne supérieure des quasi-copules dont les restrictions coïncident, mais cette borne supérieure diffère en général de la borne supérieure des prolongements d'une sous-copule. Notre travail montre que la borne supérieure des prolongements d'une sous-copule D en dimension $d > 2$ est une quasi-copule. Notre apport par rapport à la littérature existante est de proposer une démarche constructive générale pour l'obtention d'une telle borne.

Le chapitre est structuré comme suit : en premier lieu, la caractérisation de l'ensemble des prolongements d'une sous-copule à domaine de cardinalité finie est proposée, ensuite la solution en dimension 2 est présentée. Enfin, la solution en dimension 3 est établie. Un exemple de construction de la borne supérieure basé sur le lemme 4.2.1 complète le chapitre.

4.2. CARACTÉRISATION DES PROLONGEMENTS D'UNE SOUS-COPULE AVEC UN DOMAINE DE CARDINALITÉ FINIE

Avant de définir la notion de sous-copule, définissons la notion de d -croissance.

Définition 4.2.1. *Soient*

$$R_i = \{a_{i1}, a_{i2}, \dots, a_{in_i}\}, 0 = a_{i1} < a_{i2} < \dots < a_{in_i} = 1,$$

$i = 1, \dots, d$, des sous-ensembles de $[0, 1]$ de cardinalité finie, soit D une fonction réelle à d variables telle que le domaine de D soit $R_1 \times \dots \times R_d$. Soit $B = (u_1, v_1] \times \dots \times (u_d, v_d]$ un pavé dont les sommets sont situés dans $R_1 \times \dots \times R_d$, $u_1 < v_1, \dots, u_d < v_d$. La fonction D est dite d -croissante si le D -volume de B donné par

$$\begin{aligned} V_D(B) &= \sum_{t \in \mathbf{X}_{i=1}^d \{u_i, v_i\}} (-1)^{n(t)} D(t), \text{ avec} \\ n(t) &= \{k : t_k = u_k, k = 1, \dots, d\}, \end{aligned}$$

est positif ou nul. Une d -sous-copule est une fonction qui satisfait les critères suivants :

- $D : R_1 \times \dots \times R_d \rightarrow [0, 1]$ est d -croissante, R_i étant un sous-ensemble de cardinalité finie de $[0, 1]$ qui contient 0 et 1 ;
- $D(u_1, u_2, \dots, u_d) = u_i$ si $u_1 u_2 \dots u_d = u_i$, $u_i \in R_i$, $i \in \{1, \dots, d\}$.

— D est d -croissante.

Avant d'établir les bornes de l'ensemble des copules prolongements d'une sous-copule donnée D , il serait judicieux de caractériser cet ensemble. Soit D une sous-copule discrète de domaine $R_1 \times \cdots \times R_d$. Soit $J_i = \{2, \dots, n_i\}$, $i = 1, \dots, d$. Soit $p_{l_1 \dots l_d}$ défini par

$$p_{l_1 \dots l_d} = V_D((a_{1l_1-1}, a_{1l_1}] \times \cdots \times (a_{dl_d-1}, a_{dl_d}]),$$

$l_i \in J_i$, $i = 1, \dots, d$.

Lemme 4.2.1. *Toute copule C qui prolonge D se met sous la forme*

$$C(u_1, \dots, u_d) = \sum_{l_i \in J_i, i=1, \dots, d} p_{l_1 \dots l_d} C_{l_1 \dots l_d}(F_{1, l_1 \dots l_d}(u_1), \dots, F_{d, l_1 \dots l_d}(u_d)),$$

avec

$$\sum_{l_k \in J_k, k \in \{1, \dots, d\} \setminus \{i\}} p_{l_1 \dots l_d} F_{i, l_1 \dots l_d}(u_i) = u_i - a_{il_i-1}, \quad u_i \in (a_{il_i-1}, a_{il_i}],$$

$F_{i, l_1 \dots l_d}$ sont les fonctions de répartition continues concentrées sur $(a_{il_i-1}, a_{il_i}]$, $C_{l_1 \dots l_d}$ sont des copules, $l_i \in J_i$, $u_i \in [0, 1]$, $i = 1, \dots, d$.

De même, toute fonction $C : [0, 1]^d \rightarrow [0, 1]$ de la forme

$$C(u_1, \dots, u_d) = \sum_{l_i \in J_i, i=1, \dots, d} p_{l_1 \dots l_d} C_{l_1 \dots l_d}(F_{1, l_1 \dots l_d}(u_1), \dots, F_{d, l_1 \dots l_d}(u_d)),$$

est un prolongement de D avec

$$p_{l_1 \dots l_d} \neq 0,$$

$F_{i, l_1 \dots l_d}$ des fonctions de distribution continues concentrées sur $(a_{il_i-1}, a_{il_i}]$, $C_{l_1 \dots l_d}$ des copules, $l_i \in J_i$, $u_i \in [0, 1]$, $i = 1, \dots, d$.

DÉMONSTRATION. Soit une fonction C vérifiant pour $u_i \in [0, 1]$, $i = 1, \dots, d$,

$$C(u_1, \dots, u_d) = \sum_{l_i \in J_i, i=1, \dots, d} p_{l_1 \dots l_d} C_{l_1 \dots l_d}(F_{1, l_1 \dots l_d}(u_1), \dots, F_{d, l_1 \dots l_d}(u_d)),$$

avec

$$\sum_{l_k \in J_k, k \in \{1, \dots, d\} \setminus \{i\}} p_{l_1 \dots l_d} F_{i, l_1 \dots l_d}(u_i) = u_i - a_{il_i-1}, \quad u_i \in (a_{il_i-1}, a_{il_i}],$$

$l_i \in J_i$, $i = 1, \dots, d$. La fonction C est d -monotone dans tout pavé $(a_{1l_1-1}, a_{1l_1}] \times \cdots \times (a_{dl_d-1}, a_{dl_d}]$, $l_i \in J_i$, $i = 1, \dots, d$, est définie sur $[0, 1]^d$, est continue, et a des marges uniformes. De plus, elle coïncide avec D sur son domaine puisque la mesure de probabilité associée affecte la masse $p_{l_1 \dots l_d}$ au pavé $(a_{1l_1-1}, a_{1l_1}] \times \cdots \times (a_{dl_d-1}, a_{dl_d}]$, $l_i \in J_i$, $i = 1, \dots, d$. La fonction C est donc un prolongement de D . Maintenant considérons C^* un prolongement de D . Elle affecte la masse $p_{l_1 \dots l_d}$ au pavé $(a_{1l_1-1}, a_{1l_1}] \times \cdots \times (a_{dl_d-1}, a_{dl_d}]$, $l_i \in J_i$, $i = 1, \dots, d$. Considérons la

distribution conditionnelle associée à C^* dans $(a_{1l_1-1}, a_{1l_1}] \times \cdots \times (a_{dl_d-1}, a_{dl_d}]$, pour $p_{l_1 \dots l_d} > 0$, et notons la fonction de répartition associée $F_{l_1 \dots l_d}^*$, $l_i \in J_i, i = 1, \dots, d$. Pour $l_1 \in J_1, \dots, l_d \in J_d$, par application du théorème de Sklar, nous pouvons trouver une copule $C_{l_1 \dots l_d}$ et des fonctions de répartition continues concentrées sur $(a_{il_i-1}, a_{il_i}]$, $F_{i,l_1 \dots l_d}^*$, $i = 1, \dots, d$, telles que

$$F_{l_1 \dots l_d}^*(u_1, \dots, u_d) = C_{l_1 \dots l_d}(F_{1,l_1 \dots l_d}^*(u_1), \dots, F_{d,l_1 \dots l_d}^*(u_d)),$$

$u_i \in (a_{il_i-1}, a_{il_i}]$, $i = 1, \dots, d$. Le fait que C^* doit avoir des marges uniformes implique que

$$\sum_{l_k \in J_k, k \in \{1, \dots, d\} \setminus \{i\}} p_{l_1 \dots l_d} F_{i,l_1 \dots l_d}^*(u_i) = u_i - a_{il_i-1}, \quad u_i \in (a_{il_i-1}, a_{il_i}],$$

où

$$\sum_{l_i \in J_i, i=1, \dots, d} p_{l_1 \dots l_d} F_{i,l_1 \dots l_d}^*$$

représente la fonction de répartition marginale associée à C^* sur $(a_{il_i-1}, a_{il_i}]$, $i = 1, \dots, d$. \square

Pour $(u_1, \dots, u_d) \in [0, 1]^d$ fixé, trouver la valeur $UC(u_1, \dots, u_d)$ définie par

$$UC(u_1, \dots, u_d) = \sup\{C(u_1, \dots, u_d) : C \text{ est un prolongement de } D\}$$

revient à résoudre le programme

$$\sup_{\{F_{i,l_1 \dots l_d}(u_i) : l_i \in J_i, i=1, \dots, d\}} C(u_1, \dots, u_d) \quad (4.2.1)$$

avec

$$C(u_1, \dots, u_d) = \sum_{l_i \in J_i, i=1, \dots, d} p_{l_1 \dots l_d} \min(F_{1,l_1 \dots l_d}(u_1), \dots, F_{d,l_1 \dots l_d}(u_d)),$$

$$\sum_{l_k \in J_k, k \in \{1, \dots, d\} \setminus \{i\}} p_{l_1 \dots l_d} F_{i,l_1 \dots l_d}(u_i) = u_i - a_{il_i-1}, \quad u_i \in (a_{il_i-1}, a_{il_i}],$$

$$0 \leq F_{i,l_1 \dots l_d}(u_i) \leq 1$$

$l_i \in J_i$, $i = 1, \dots, d$, où les copules $C_{l_1 \dots l_d}$ sont prises égales à la borne supérieure de Fréchet :

$$\begin{aligned} [0, 1]^d &\rightarrow [0, 1] \\ (u_1, \dots, u_d) &\mapsto \min(u_1, \dots, u_d), \end{aligned}$$

$l_1 \in J_1, \dots, l_d \in J_d$, $i = 1, \dots, d$. Ceci est vrai car pour $F_{i,l_1 \dots l_d}$ fixés, $l_1 \in J_1, \dots, l_d \in J_d$, $i = 1, \dots, d$, le prolongement pour lequel les copules $C_{l_1 \dots l_d}$ sont prises égales à la borne supérieure de Fréchet est une borne supérieure. Les contraintes du programme émanent du fait que toute copule a des marges uniformes et elles se mettent sous la forme :

$$\sum_{l_i=j_i, l_k \leq t_k, k \neq i} p_{l_1 \dots l_d} F_{i,l_1 \dots l_d} \leq \max(0, \min(I d - a_{ij_i-1}, \sum_{l_i=j_i, l_k \leq t_k, k \neq i} p_{l_1 \dots l_d})) \quad (4.2.2)$$

où Id est mis pour l'application identité, $Id(u) = u$ pour tout $u \in [0, 1]$, $j_i, j_i - 1 \in J_i$, $t_k \in J_k$, $k \neq i$, $i = 1, \dots, d$.

Lemme 4.2.2. *En tout point $(u_1, \dots, u_d) \in [0, 1]^d$, il existe une copule (dépendant du point (u_1, \dots, u_d)) qui étend la sous-copule D et qui atteint la borne supérieure (4.2.1).*

DÉMONSTRATION. Pour $(u_1, \dots, u_d) \in [0, 1]^d$ fixé, considérons l'ensemble

$$E = \{C(u_1, \dots, u_d) : C \text{ est un prolongement de } D\}.$$

Cet ensemble est un sous-ensemble de $[0, 1]$. Si on considère la borne supérieure de cet ensemble $\sup E$, on peut trouver une suite $(C_n(u_1, \dots, u_d))_{n>0}$ dont la limite est $\sup E$. Considérons la suite des prolongements $(C_n)_{n>0}$. Cette suite est uniformément bornée et équicontinue; il existe une sous-suite $(C_{n_k})_{n_k>0}$ qui converge uniformément. La limite de cette sous-suite est une copule qui atteint la borne supérieure en (u_1, \dots, u_d) . \square

Le dernier résultat du paragraphe montre que la borne supérieure est une quasi-copule.

Lemme 4.2.3. *La fonction UC définie par*

$$\begin{aligned} UC : [0, 1]^d &\rightarrow [0, 1] \\ (u_1, \dots, u_d) &\mapsto \sup\{C(u_1, \dots, u_d), C \text{ est un prolongement de } D\} \end{aligned}$$

est une quasi-copule.

DÉMONSTRATION. La fonction UC a des marges uniformes car pour $u_1 \in [0, 1]$, $UC(u_1, 1, \dots, 1) = \sup\{C(u_1, 1, \dots, 1) = u_1, C \text{ est un prolongement de } D\}$. La fonction UC est monotone en chaque coordonnée : pour $0 \leq u_1 \leq u'_1 \leq 1$, $u_i \in [0, 1]$, $i = 2, \dots, d$ fixés, nous avons

$$\begin{aligned} UC(u_1, \dots, u_d) &= \sup\{C(u_1, \dots, u_d), C \text{ est un prolongement de } D\} \\ &\leq \sup\{C(u'_1, u_2, \dots, u_d), C \text{ est un prolongement de } D\} \\ &= UC(u'_1, u_2, \dots, u_d). \end{aligned}$$

Aussi, UC est nulle dès qu'une des coordonnées u_1, \dots, u_d est nulle parce que UC correspond dans ce cas au supremum d'un ensemble réduit à 0. La fonction UC coïncide avec D sur son domaine. En effet pour $(u_1, u_2, \dots, u_d) \in R_1 \times \dots \times R_d$, l'ensemble $\{C(u_1, \dots, u_d), C \text{ est un prolongement de } D\}$ est réduit au singleton $\{D(u_1, u_2, \dots, u_d)\}$ et on a

$$UC(u_1, \dots, u_d) = D(u_1, \dots, u_d).$$

\square

4.3. LA SOLUTION EN DIMENSION 2

Dans le cas de la dimension 2, pour $(u_1, u_2) \in (a_{1j_1-1}, a_{1j_1}] \times (a_{2j_2-1}, a_{2j_2}]$, $j_1 - 1, j_1 \in J_1, j_2 - 1, j_2 \in J_2$, nous avons à résoudre

$$\sup_{\{F_{i,j_1j_2}(u_i): j_i \in J_i, i=1,2\}} C(u_1, u_2) \quad (4.3.1)$$

avec

$$\begin{aligned} C(u_1, u_2) = & D(a_{1j_1-1}, a_{2j_2-1}) + \sum_{l_2 \leq j_2-1} p_{j_1l_2} F_{1,j_1l_2}(u_1) + \\ & \sum_{l_1 \leq j_1-1} p_{l_1j_2} F_{2,l_1j_2}(u_2) + p_{j_1j_2} \min(F_{1,j_1j_2}(u_1), F_{2,j_1j_2}(u_2)) \end{aligned}$$

sous contraintes

$$\begin{aligned} \sum_{l_2 \leq t_2} p_{j_1l_2} F_{1,j_1l_2}(u_1) & \leq \min(u_1 - a_{1j_1-1}, \sum_{l_2 \leq t_2} p_{j_1l_2}), \\ \sum_{l_1 \leq t_1} p_{l_1j_2} F_{2,l_1j_2}(u_2) & \leq \min(u_2 - a_{2j_2-1}, \sum_{l_1 \leq t_1} p_{l_1j_2}), \\ 0 \leq F_{1,j_1l_2}(u_1) & \leq 1, \\ 0 \leq F_{2,l_1j_2}(u_2) & \leq 1, \end{aligned} \quad (4.3.2)$$

$l_1 \leq j_1, l_2 \leq j_2, t_1 \in \{j_1 - 1, j_1\}, t_2 \in \{j_2 - 1, j_2\}, j_1 - 1, j_1 \in J_1, j_2 - 1, j_2 \in J_2$.

La démarche de résolution consistera à trouver dans un premier temps un majorant de l'ensemble des prolongements puis à prouver que ce majorant est un prolongement.

Théorème 4.3.1. *En $(u_1, u_2) \in (a_{1j_1-1}, a_{1j_1}] \times (a_{2j_2-1}, a_{2j_2}]$, la borne supérieure est*

$$UC(u_1, u_2) = D(a_{1j_1-1}, a_{2j_2-1}) + \min(f_{11} + f_{22}, f_{12} + f_{21})$$

avec

$$\begin{aligned} f_{11} &= \min(u_1 - a_{1j_1-1}, \sum_{l_2 \leq j_2-1} p_{j_1l_2}); \\ f_{22} &= \min(u_2 - a_{2j_2-1}, \sum_{l_1 \leq j_1} p_{l_1j_2}); \\ f_{12} &= \min(u_1 - a_{1j_1-1}, \sum_{l_2 \leq j_2} p_{j_1l_2}); \\ f_{21} &= \min(u_2 - a_{2j_2-1}, \sum_{l_1 \leq j_1-1} p_{l_1j_2}). \end{aligned}$$

$j_1 - 1, j_1 \in J_1, j_2 - 1, j_2 \in J_2$.

DÉMONSTRATION. Soit C un prolongement de la sous-copule D ; en $(u_1, u_2) \in (a_{1j_1-1}, a_{1j_1}] \times (a_{2j_2-1}, a_{2j_2}]$, nous avons

$$\begin{aligned} C(u_1, u_2) = & D(a_{1j_1-1}, a_{2j_2-1}) + \min\left(\sum_{l_2 \leq j_2} p_{j_1l_2} F_{1,j_1l_2}(u_1) + \sum_{l_1 \leq j_1-1} p_{l_1j_2} F_{2,l_1j_2}(u_2), \right. \\ & \left. \sum_{l_2 \leq j_2-1} p_{j_1l_2} F_{1,j_1l_2}(u_1) + \sum_{l_1 \leq j_1} p_{l_1j_2} F_{2,l_1j_2}(u_2)\right), \end{aligned}$$

$j_1 - 1, j_1 \in J_1, j_2 - 1, j_2 \in J_2$. À partir des inégalités

$$\begin{aligned} \sum_{l_2 \leq j_2} p_{j_1 l_2} F_{1, j_1 l_2}(u_1) &\leq \min(u_1 - a_{1j_1-1}, \sum_{l_2 \leq j_2} p_{j_1 l_2}) \\ \sum_{l_2 \leq j_2-1} p_{j_1 l_2} F_{1, j_1 l_2}(u_1) &\leq \min(u_1 - a_{1j_1-1}, \sum_{l_2 \leq j_2-1} p_{j_1 l_2}) \\ \sum_{l_1 \leq j_1} p_{l_1 j_2} F_{2, l_1 j_2}(u_2) &\leq \min(u_2 - a_{2j_2-1}, \sum_{l_1 \leq j_1} p_{l_1 j_2}) \\ \sum_{l_1 \leq j_1-1} p_{l_1 j_2} F_{2, l_1 j_2}(u_2) &\leq \min(u_2 - a_{2j_2-1}, \sum_{l_1 \leq j_1-1} p_{l_1 j_2}) \end{aligned}$$

nous pouvons déduire que $C(u_1, u_2) \leq UC(u_1, u_2)$. Le lemme suivant prouve qu'il existe un prolongement C telle que $C(u_1, u_2) = UC(u_1, u_2)$; ce qui prouve que $UC(u_1, u_2)$ est la borne supérieure en (u_1, u_2) . \square

Lemme 4.3.1. *Soit $(u_1, u_2) \in (a_{1j_1-1}, a_{1j_1}] \times (a_{2j_2-1}, a_{2j_2}]$, $j_1 - 1, j_1 \in J_1, j_2 - 1, j_2 \in J_2$, il existe un prolongement qui atteint la borne UC en (u_1, u_2) . En particulier, le prolongement C vérifiant :*

$$\begin{aligned} p_{j_1 j_2} F_{1, j_1 j_2} &= \max(0, \min(Id - a_{1j_1-1} - \sum_{l_2 \leq t_2} p_{j_1 l_2} + p_{j_1 j_2}, p_{j_1 j_2})) \\ p_{j_1 j_2} F_{2, j_1 j_2} &= \max(0, \min(Id - a_{2j_2-1} - \sum_{l_1 \leq t_1} p_{l_1 j_2} + p_{j_1 j_2}, p_{j_1 j_2})) \end{aligned}$$

$p_{j_1 j_2} > 0, j_1 - 1, j_1 \in J_1, j_2 - 1, j_2 \in J_2$ satisfait la relation $C(u_1, u_2) = UC(u_1, u_2)$.

DÉMONSTRATION. Soit le prolongement C avec les égalités suivantes :

$$\begin{aligned} \sum_{l_2 \leq j_2} p_{j_1 l_2} F_{1, j_1 l_2} &= \max(0, \min(Id - a_{1j_1-1}, \sum_{l_2 \leq j_2} p_{j_1 l_2})), \\ \sum_{l_1 \leq j_1} p_{l_1 j_2} F_{2, l_1 j_2} &= \max(0, \min(u_2 - a_{2j_2-1}, \sum_{l_1 \leq j_1} p_{l_1 j_2})), \end{aligned} \quad (4.3.3)$$

et

$$\begin{aligned} \sum_{l_2 \leq j_2-1} p_{j_1 l_2} F_{1, j_1 l_2} &= \max(0, \min(Id - a_{1j_1-1}, \sum_{l_2 \leq j_2-1} p_{j_1 l_2})), \\ \sum_{l_1 \leq j_1-1} p_{l_1 j_2} F_{2, l_1 j_2} &= \max(0, \min(Id - a_{2j_2-1}, \sum_{l_1 \leq j_1-1} p_{l_1 j_2})); \end{aligned} \quad (4.3.4)$$

$j_1 \in J_1, j_2 \in J_2$. Les seules fonctions capables de vérifier ces égalités (4.3.3), (4.3.4), pour tous $(j_1, j_2) \in J_1 \times J_2$ sont ceux du lemme. Les fonctions $F_{1, j_1 j_2}, F_{2, j_1 j_2}, j_1 \in J_1, j_2 \in J_2$ sont des fonctions de répartition et C est un prolongement de la sous-copule D . En outre, les égalités (4.3.3), (4.3.4) nous permettent de dire que $C(u_1, u_2) = UC(u_1, u_2), (u_1, u_2) \in (a_{1j_1-1}, a_{1j_1}] \times (a_{2j_2-1}, a_{2j_2}]$. \square

4.4. LA BORNE SUPÉRIEURE EN DIMENSION 3

Soit D une sous-copule de domaine $R_1 \times R_2 \times R_3$, $R_i = \{a_{i1}, a_{i2}, \dots, a_{in_i}\}$, $0 = a_{i1} < a_{i2} < \dots < a_{in_i} = 1, i = 1, 2, 3$. Soit $(u_1, u_2, u_3) \in (a_{1j_1-1}, a_{1j_1}] \times (a_{2j_2-1}, a_{2j_2}] \times (a_{3j_3-1}, a_{3j_3}]$ fixé. Soit

$$R_i^* = R_i \cup \{u_i\} = \{b_{i2}, \dots, b_{in_i+1}\},$$

$$b_{i1} = a_{i1}, \dots, b_{ij_i-1} = a_{ij_i-1}, b_{ij_i} = u_i, b_{ij_i+1} = a_{ij_i}, \dots, b_{in_i+1} = a_{in_i}, i = 1, 2, 3.$$

Soit D' un prolongement de D sur $R_1^* \times R_2^* \times R_3^*$. Posons

$$p'_{l_1 l_2 l_3} = V_{D'}((b_{1l_1-1}, b_{1l_1}] \times (b_{2l_2-1}, b_{2l_2}] \times (b_{3l_3-1}, b_{3l_3}]),$$

$J_i^* = \{2, \dots, n_i + 1\}$, $i = 1, 2, 3$, avec les égalités suivantes

$$\begin{aligned} p'_{l_1 l_2 l_3} &= p_{l'_1 l'_2 l'_3}, l'_i = l_i \text{ si } l_i < j_i, l'_i = l_i - 1 \text{ sinon, } i = 1, 2, 3, \\ (l_1, l_2, l_3) &\in J_1^* \setminus \{j_1, j_1 + 1\} \times J_2^* \setminus \{j_2, j_2 + 1\} \times J_3^* \setminus \{j_3, j_3 + 1\}. \end{aligned}$$

Il reste $n_1 \times n_2 \times n_3 - (n_1 - 2) \times (n_2 - 2) \times (n_3 - 2)$ quantités à définir pour définir complètement D' . Ces quantités sont :

$$\begin{aligned} p'_{l_1 l_2 l_3} &\geq 0, (l_1, l_2, l_3) \in \{j_1, j_1 + 1\} \times J_2^* \times J_3^*, \\ p'_{l_1 j_2 l_3} &\geq 0, (l_1, l_2, l_3) \in J_1^* \setminus \{j_1, j_1 + 1\} \times \{j_2, j_2 + 1\} \times J_3^*, \\ p'_{l_1 l_2 j_3} &\geq 0, (l_1, l_2, l_3) \in J_1^* \setminus \{j_1, j_1 + 1\} \times J_2^* \setminus \{j_2, j_2 + 1\} \times \{j_3, j_3 + 1\}. \end{aligned}$$

Elles sont définies de la manière suivante :

$$\begin{aligned} p'_{j_1+1 l_2 l_3} + p'_{j_1 l_2 l_3} &= p_{j_1 l_2 l_3}, l_2 \in J_2^* \setminus \{j_2, j_2 + 1\}, l_3 \in J_3^* \setminus \{j_3, j_3 + 1\}, \\ p'_{l_1 j_2+1 l_3} + p'_{l_1 j_2 l_3} &= p_{l_1 j_2 l_3}, l_1 \in J_1^* \setminus \{j_1, j_1 + 1\}, l_3 \in J_3^* \setminus \{j_3, j_3 + 1\}, \\ p'_{l_1 l_2 j_3+1} + p'_{l_1 l_2 j_3} &= p_{l_1 l_2 j_3}, l_1 \in J_1^* \setminus \{j_1, j_1 + 1\}, l_2 \in J_2^* \setminus \{j_2, j_2 + 1\}, \\ p'_{j_1 l_2 j_3} + p'_{j_1+1 l_2 j_3} + p'_{j_1 l_2 j_3+1} + p'_{j_1+1 l_2 j_3+1} &= p_{j_1 l_2 j_3}, l_2 \in J_2^* \setminus \{j_2, j_2 + 1\}, \\ p'_{j_1 j_2 l_3} + p'_{j_1+1 j_2 l_3} + p'_{j_1 j_2+1 l_3} + p'_{j_1+1 j_2+1 l_3} &= p_{j_1 j_2 l_3}, l_3 \in J_3^* \setminus \{j_3, j_3 + 1\}, \\ p'_{l_1 j_2 j_3} + p'_{l_1 j_2+1 j_3} + p'_{l_1 j_2 j_3+1} + p'_{l_1 j_2+1 j_3+1} &= p_{l_1 j_2 j_3}, l_1 \in J_1^* \setminus \{j_1, j_1 + 1\}, \\ p'_{j_1 j_2 j_3} + p'_{j_1+1 j_2 j_3} + p'_{j_1 j_2+1 j_3} + p'_{j_1+1 j_2+1 j_3} + p'_{j_1 j_2 j_3+1} + p'_{j_1+1 j_2 j_3+1} &+ p'_{j_1+1 j_2 j_3+1} + p'_{j_1+1 j_2+1 j_3+1} = p_{j_1 j_2 j_3}, \\ \sum_{l_2 \in J_2^*, l_3 \in J_3^*} p'_{j_1 l_2 l_3} &= u_1 - a_{1j_1-1}, \\ \sum_{l_1 \in J_1^*, l_3 \in J_3^*} p'_{l_1 j_2 l_3} &= u_2 - a_{2j_2-1}, \\ \sum_{l_1 \in J_1^*, l_2 \in J_2^*} p'_{l_1 l_2 j_3} &= u_3 - a_{3j_3-1}. \end{aligned} \tag{4.4.1}$$

Rechercher la borne supérieure des prolongements en (u_1, u_2, u_3) revient à résoudre le programme suivant :

$$\sup_{\{p'_{l_1 l_2 j_3} : l_i \in J_i^*, i=1,2,3\}} D(a_{1j_1-1}, a_{2j_2-1}, a_{3j_3-1}) + \sum_{\{l_i \leq j_i, i=1,2,3\} \cap \{\bigcup_{i=1,2,3} \{l_i = j_i\}\}} p'_{l_1 l_2 l_3}$$

sous contraintes

$$\begin{aligned} 0 &\leq p'_{l_1 l_2 l_3} \leq p_{l_1 l_2 l_3}, l_i \leq j_i, i = 1, 2, 3, \\ \sum_{l_2 \leq j_2, l_3 \leq j_3} p'_{j_1 l_2 l_3} &\leq u_1 - a_{1j_1-1}, \\ \sum_{l_1 \leq j_1, l_3 \leq j_3} p'_{l_1 j_2 l_3} &\leq u_2 - a_{2j_2-1}, \\ \sum_{l_1 \leq j_1, l_2 \leq j_2} p'_{l_1 l_2 j_3} &\leq u_3 - a_{3j_3-1}. \end{aligned}$$

Lemme 4.4.1. *Problème réduit. Posons*

$$\begin{aligned} S_k &= \{0, 1\} \cup \{a_{kj_k-1}, a_{kj_k}\}, k = 1, 2, 3, \\ S_k^* &= S_k \cup \{u_k\}, k = 1, 2, 3. \end{aligned}$$

Soit D_0 la sous-copule restriction de D sur $S_1 \times S_2 \times S_3$;

- (a) si D' prolonge D sur $R_1^* \times R_2^* \times R_3^*$, alors il existe D'_0 un prolongement de D_0 sur $S_1^* \times S_2^* \times S_3^*$ tel que $D'(u_1, u_2, u_3) = D'_0(u_1, u_2, u_3)$;
- (b) si D'_0 prolonge D_0 sur $S_1^* \times S_2^* \times S_3^*$, alors il existe D' un prolongement de D sur $R_1^* \times R_2^* \times R_3^*$ tel que $D'_0(u_1, u_2, u_3) = D'(u_1, u_2, u_3)$.

DÉMONSTRATION. Nous ferons la démonstration en utilisant les mesures induites par les sous-copules.

- (a) Pour le premier résultat, il suffit de poser D'_0 comme la restriction de D' sur $S_1^* \times S_2^* \times S_3^*$. Le domaine de D_0 étant inclus dans celui de D , D' est aussi un prolongement de D_0 ; par conséquent D'_0 est un prolongement de D_0 .
- (b) Soient P la mesure de probabilité induite par D , Q la mesure de probabilité induite par D_0 , Q' la mesure de probabilité induite par D'_0 . Posons

$$\begin{aligned} 0 &= c_{i0}, c_{i1} = a_{ij_i-1}, c_{i2} = a_{ij_i}, c_{i3} = 1, i = 1, 2, 3; \\ 0 &= d_{i0}, d_{i1} = a_{ij_i-1}, d_{i2} = u_i, d_{i3} = a_{ij_i}, d_{i4} = 1, i = 1, 2, 3. \end{aligned}$$

$$\begin{aligned} p_{l_1 l_2 l_3} &= P((a_{1l_1-1}, a_{1l_1}] \times (a_{2l_2-1}, a_{2l_2}] \times (a_{3l_3-1}, a_{3l_3}]), \\ q_{l_1 l_2 l_3} &= Q((c_{1l_1-1}, c_{1l_1}] \times (c_{2l_2-1}, c_{2l_2}] \times (c_{3l_3-1}, c_{3l_3}]), \quad 1 \leq l_1, l_2, l_3 \leq 3, \\ q'_{l_1 l_2 l_3} &= Q'((d_{1l_1-1}, d_{1l_1}] \times (d_{2l_2-1}, d_{2l_2}] \times (d_{3l_3-1}, d_{3l_3}]), \quad 1 \leq l_1, l_2, l_3 \leq 4. \end{aligned}$$

Nous posons

$$\begin{aligned}
p_{l_1 l_2 l_3}^* &= p_{l_1 - I(l_1 > j_1) l_2 - I(l_2 > j_2) l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_k \notin \{j_k, j_k + 1\}, k = 1, 2, 3, \\
p_{j_1 l_2 l_3}^* &= \frac{q'_{21+3I(l_2 > j_2) 1+3I(l_3 > j_3)}}{q_{21+2I(l_2 > j_2) 1+2I(l_3 > j_3)}} p_{l_1 - I(l_1 > j_1) l_2 - I(l_2 > j_2) l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_k \notin \{j_k, j_k + 1\}, k = 2, 3 \\
p_{j_1+1 l_2 l_3}^* &= \frac{q'_{3 \ 1+3I(l_2 > j_2) \ 1+3I(l_3 > j_3)}}{q_{2 \ 1+2I(l_2 > j_2) \ 1+2I(l_3 > j_3)}} p_{j_1 \ l_2 - I(l_2 > j_2) \ l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_k \notin \{j_k, j_k + 1\}, k = 2, 3 \\
p_{j_1 j_2 l_3}^* &= \frac{q'_{22 \ 1+3I(l_3 > j_3)}}{q_{22 \ 1+2I(l_3 > j_3)}} p_{j_1 j_2 \ l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_3 \notin \{j_3, j_3 + 1\}, \\
p_{j_1+1 j_2 l_3}^* &= \frac{q'_{32 \ 1+3I(l_3 > j_3)}}{q_{22 \ 1+2I(l_3 > j_3)}} p_{j_1 j_2 \ l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_3 \notin \{j_3, j_3 + 1\}, \\
p_{j_1 j_2+1 l_3}^* &= \frac{q'_{32 \ 1+3I(l_3 > j_3)}}{q_{22 \ 1+2I(l_3 > j_3)}} p_{j_1 j_2 \ l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_3 \notin \{j_3, j_3 + 1\}, \\
p_{j_1+1 j_2+1 l_3}^* &= \frac{q'_{33 \ 1+3I(l_3 > j_3)}}{q_{22 \ 1+2I(l_3 > j_3)}} p_{j_1 j_2 \ l_3 - I(l_3 > j_3)} \\
&\quad \text{si } l_3 \notin \{j_3, j_3 + 1\}, \\
p_{j_1 j_2 j_3}^* &= \frac{q'_{222}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1 j_2+1 j_3}^* &= \frac{q'_{322}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1+1 j_2 j_3}^* &= \frac{q'_{322}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1+1 j_2+1 j_3}^* &= \frac{q'_{332}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1 j_2 j_3+1}^* &= \frac{q'_{223}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1 j_2+1 j_3+1}^* &= \frac{q'_{323}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1+1 j_2 j_3+1}^* &= \frac{q'_{323}}{q_{222}} p_{j_1 j_2 j_3}, \\
p_{j_1+1 j_2+1 j_3+1}^* &= \frac{q'_{333}}{q_{222}} p_{j_1 j_2 j_3},
\end{aligned}$$

$1 \leq l_i \leq n_i, i = 1, 2, 3$. Les formules de $p_{l_1 j_2 j_3}^*, p_{l_1 j_2+1 j_3}^*, p_{l_1 j_2 j_3+1}^*, p_{l_1 j_2+1 j_3+1}^*, p_{j_1 l_2 j_3}^*, p_{j_1+1 l_2 j_3}^*, p_{j_1 l_2 j_3+1}^*, p_{j_1+1 l_2 j_3+1}^*$ sont obtenues par symétrie des formules $p_{j_1 j_2 l_3}^*, p_{j_1 j_2+1 l_3}^*, p_{j_1+1 j_2 l_3}^*, p_{j_1+1 j_2+1 l_3}^*, 1 \leq l_i \leq n_i, i = 1, 2, 3$. Définissons la loi de probabilité P^* sur $R_1^* \times R_2^* \times R_3^*$:

$$P^*((a_{1l_1-1}, a_{1l_1}] \times (a_{2l_2-1}, a_{2l_2}] \times (a_{3l_3-1}, a_{3l_3}]) = p_{l_1 l_2 l_3}^*,$$

$1 \leq l_i \leq n_i, i = 1, 2, 3$. La loi de probabilité P^* a des marges uniformes. Soit D' la sous-copule associée à P^* . Nous avons $D'(u_1, u_2, u_3) = D'_0(u_1, u_2, u_3)$ et D' est un prolongement de D .

□

Pour la suite, nous utilisons la sous-copule D_0 restriction de D sur $S_1 \times S_2 \times S_3$.

Posons $\mathcal{D} = \{D'_0 : D'_0 \text{ est un prolongement de } D_0 \text{ sur } S_1^* \times S_2^* \times S_3^*\}$. Posons

$$\begin{aligned} L &= \{(l_1, l_2, l_3) \in \{1, 2, 3\}^3 : \prod_{i=1}^3 (2 - l_i) = 0\}, \\ L' &= \{(l_1, l_2, l_3) \in \{1, 2, 3, 4\}^3 : \prod_{i=1}^3 (2 - l_i) = 0\}, \\ L_0 &= \{(l_1, l_2, l_3) \in L : l_i \leq 2, i = 1, 2, 3\}, \\ L'_0 &= \{(l_1, l_2, l_3) \in L' : l_i \leq 2, i = 1, 2, 3\}, \\ L_k &= \{(l_1, l_2, l_3) \in L : l_k = 2\}, \\ L'_k &= \{(l_1, l_2, l_3) \in L' : l_k = 2\}, \\ L_{0k} &= L_0 \cap L_k, \\ L'_{0k} &= L'_0 \cap L'_k. \end{aligned}$$

Soit Q' la mesure de probabilité associée à D'_0 , posons

$$0 = d_{i0}, d_{i1} = a_{ij_i-1}, d_{i2} = u_i, d_{i3} = a_{ij_i}, d_{i4} = 1, i = 1, 2, 3.$$

$$q'_{l_1 l_2 l_3} = Q'((d_{1l_1-1}, d_{1l_1}] \times (d_{2l_2-1}, d_{2l_2}] \times (d_{3l_3-1}, d_{3l_3}]),$$

$1 \leq l_1, l_2, l_3 \leq 4$. Nous avons

$$D'_0(u_1, u_2, u_3) = D(a_{1j_1-1}, a_{2j_2-1}, a_{3j_3-1}) + \sum_{(l_1, l_2, l_3) \in L'_0} q'_{l_1 l_2 l_3}.$$

Posons

$$\mathbb{O}' = \sum_{(l_1, l_2, l_3) \in L'_0} q'_{l_1 l_2 l_3}.$$

On cherche à maximiser \mathbb{O}' sous contraintes

$$\begin{aligned} q'_l &\geq 0 \quad l \in L, \\ q'_{2l_2 l_3} &\leq q_2 \, l_2 - I(l_2 > 2) \, l_3 - I(l_3 > 2), \, l_2 \notin \{2, 3\}, \, l_3 \notin \{2, 3\}, \, (2, l_2, l_3) \in L', \\ q'_{l_1 2l_3} &\leq q_{l_1 - I(l_1 > 2)} \, 2 \, l_3 - I(l_3 > 2), \, l_1 \notin \{2, 3\}, \, l_3 \notin \{2, 3\}, \, (l_1, 2, l_3) \in L', \\ q'_{l_1 l_2 2} &\leq q_{l_1 - I(l_1 > 2)} \, l_2 - I(l_2 > 2) \, 2, \, l_1 \notin \{2, 3\}, \, l_2 \notin \{2, 3\}, \, (l_1, l_2, 2) \in L', \\ q'_{22l_3} + q'_{23l_3} + q'_{32l_3} &\leq q_{22} \, l_3 - I(l_3 > 2), \, l_3 \notin \{2, 3\}, \, (2, 2, l_3) \in L', \\ q'_{2l_2 2} + q'_{2l_2 3} + q'_{3l_2 2} &\leq q_{22} \, l_3 - I(l_3 > 2), \, l_2 \notin \{2, 3\}, \, (2, l_2, 2) \in L', \\ q'_{l_1 22} + q'_{l_1 23} + q'_{l_1 32} &\leq q_{l_1 - I(l_1 > 2)} \, 22, \, l_1 \notin \{2, 3\}, \, (l_1, 2, 2) \in L', \\ q'_{222} + q'_{223} + q'_{232} + q'_{233} + q'_{322} + q'_{323} + q'_{332} &\leq q_{222}. \end{aligned}$$

On pose

$$\begin{aligned}
M_k &= u_k - a_{kj_k-1}, k = 1, 2, 3, \\
\pi_{211} &= \min(M_1, q_{211}), \\
\pi_{121} &= \min(M_2, q_{121}), \\
\pi_{112} &= \min(M_3, q_{112}), \\
M_1^c &= M_1 - \pi_{211}, \\
M_2^c &= M_2 - \pi_{121}, \\
M_3^c &= M_3 - \pi_{112}, \\
\pi_{221} &= \min(M_1^c, M_2^c, q_{211}), \\
\pi_{212} &= \min(M_1^c, M_3^c, q_{212}), \\
\pi_{122} &= \min(M_2^c, M_3^c, q_{122}), \\
\mu_1 &= \min(M_1^c, \pi_{221} + \pi_{212}), \\
\mu_2 &= \min(M_2^c, \pi_{221} + \pi_{122}), \\
\mu_3 &= \min(M_3^c, \pi_{212} + \pi_{122}), \\
\pi_{222} &= \min(M_1^c - \mu_1, M_2^c - \mu_2, M_3^c - \mu_3, q_{222}).
\end{aligned}$$

On peut montrer qu'on peut poser sans perte de généralité

$$\begin{aligned}
q'_{211} &= \pi_{211}, \\
q'_{121} &= \pi_{121}, \\
q'_{112} &= \pi_{112}.
\end{aligned}$$

Théorème 4.4.1. *Résultat principal*

$$\begin{aligned}
\sup\{D'_0(u_1, u_2, u_3) : D'_0 \in \mathcal{D}\} &= q_{111} + \pi_{211} + \pi_{121} + \pi_{112} + \pi_{222} + \\
&\quad \min\left\{\frac{\mu_1 + \mu_2 + \mu_3}{2}, \mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}\right\}
\end{aligned}$$

DÉMONSTRATION. La démonstration sera structurée en étapes.

(Étape 1) Montrons qu'on peut poser sans perte de généralité

$$\begin{aligned}
q'_{211} &= \pi_{211}, \\
q'_{121} &= \pi_{121}, \\
q'_{112} &= \pi_{112}.
\end{aligned}$$

Soit q'_l , $l \in L'$ un candidat qui satisfait toutes les contraintes. Supposons que $q'_{211} < \pi_{211}$. On pose

$$\begin{aligned}
q''_{211} &= \pi_{211}, \\
q''_{311} &= \pi_{211} - q'_{211}, \\
q''_{2l_2l_3} &= \frac{M_1^c - \pi_{211}}{M_1^c - q'_{211}} q'_{2l_2l_3}, (2, l_2, l_3) \in L', (l_2, l_3) \neq (1, 1), \\
q''_{3l_2l_3} &= q_{3l_2l_3} - q'_{3l_2l_3}, \\
q''_l &= q'_l \text{ si } l_1 \neq 2, 3, l \in L'.
\end{aligned}$$

On trouve que q_l'' , $l \in L'$ satisfait toutes les contraintes. De plus,

$$\sum_{l \in L'_0} q_l'' - q_l' = \sum_{l \in L'_{01}} q_l'' - q_l' \geq 0$$

À partir de maintenant, nous poserons

$$\begin{aligned} q'_{211} &= \pi_{211}, \\ q'_{121} &= \pi_{121}, \\ q'_{112} &= \pi_{112}. \end{aligned}$$

(Étape 2) Montrons que si

$$\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) < q_{221} + q_{212} + q_{122}$$

alors il existe $k \in \{1, 2, 3\}$ tel que

$$\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) = M_k^c + \pi_{2-I_1(k)} - \pi_{2-I_2(k)} - \pi_{2-I_3(k)}.$$

Supposons sans perte de généralité que $M_1^c \leq M_2^c \leq M_3^c$.

(1) Si $M_1^c \leq \min(q_{221}, q_{212})$, $M_2^c \leq q_{122}$, nous avons

$$\begin{aligned} \pi_{221} &= M_1^c, \\ \pi_{212} &= M_1^c, \\ \pi_{122} &= M_2^c, \\ \mu_1 &= M_1^c, \\ \mu_2 &= M_2^c, \\ \mu_3 &= \min(M_3^c, \pi_{212} + \pi_{122}), \\ \min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= M_1^c + M_2^c \end{aligned}$$

(2) Si $M_1^c \leq \min(q_{221}, q_{212})$, $M_2^c > q_{122}$, nous avons

$$\begin{aligned} \pi_{221} &= M_1^c, \\ \pi_{212} &= M_1^c, \\ \pi_{122} &= q_{122}, \\ \mu_1 &= M_1^c, \\ \mu_2 &= \min(M_2^c, M_1^c + q_{122}), \\ \mu_3 &= \min(M_3^c, M_1^c + q_{122}), \\ \min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= M_1^c + q_{122} \end{aligned}$$

(3) Si $q_{221} < M_1^c \leq q_{212}$, $M_2^c < q_{122}$, nous avons

$$\begin{aligned}
\pi_{221} &= q_{221}, \\
\pi_{212} &= M_1^c, \\
\pi_{122} &= M_2^c, \\
\mu_1 &= M_1^c, \\
\mu_2 &= M_2^c, \\
\mu_3 &= \min(M_3^c, M_1^c + M_2^c), \\
\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= \min(M_3^c + q_{221}, M_1^c + M_2^c)
\end{aligned}$$

(4) Si $q_{221} < M_1^c \leq q_{212}$, $M_2^c > q_{122}$, nous avons

$$\begin{aligned}
\pi_{221} &= q_{221}, \\
\pi_{212} &= M_1^c, \\
\pi_{122} &= q_{122}, \\
\mu_1 &= M_1^c, \\
\mu_2 &= \min(M_2^c, q_{122} + q_{221}), \\
\mu_3 &= \min(M_3^c, M_1^c + q_{122}), \\
\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= M_1^c + q_{122}
\end{aligned}$$

(5) Si $q_{212} < M_1^c \leq q_{221}$, $M_2^c \leq q_{122}$, nous avons

$$\begin{aligned}
\pi_{221} &= M_1^c, \\
\pi_{212} &= q_{212}, \\
\pi_{122} &= M_2^c, \\
\mu_1 &= M_1^c, \\
\mu_2 &= M_1^c, \\
\mu_3 &= \min(M_3^c, M_2^c + q_{212}), \\
\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= M_2^c + q_{212}
\end{aligned}$$

(6) Si $q_{212} < M_1^c \leq q_{221}$, $M_2^c > q_{122}$, nous avons

$$\begin{aligned}
\pi_{221} &= M_1^c, \\
\pi_{212} &= q_{212}, \\
\pi_{122} &= q_{122}, \\
\mu_1 &= M_1^c, \\
\mu_2 &= \min(M_2^c, M_1^c + q_{122}), \\
\mu_3 &= \min(M_3^c, M_1^c + q_{212}), \\
\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= \min(M_1^c + q_{122}, M_2^c + q_{212})
\end{aligned}$$

(7) Si $\max(q_{212}, q_{221}) < M_1^c$, $M_2^c \leq q_{122}$, nous avons

$$\begin{aligned}
\pi_{221} &= q_{221}, \\
\pi_{212} &= q_{212}, \\
\pi_{122} &= M_2^c, \\
\mu_1 &= \min(M_1^c, q_{221} + q_{212}), \\
\mu_2 &= M_2^c, \\
\mu_3 &= \min(M_3^c, M_2^c + q_{212}), \\
\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= \min(M_2^c + q_{212}, M_3^c + q_{221})
\end{aligned}$$

(8) Si $\max(q_{212}, q_{221}) < M_1^c$, $M_2^c > q_{122}$, nous avons

$$\begin{aligned}
\pi_{221} &= q_{221}, \\
\pi_{212} &= q_{212}, \\
\pi_{122} &= q_{122}, \\
\mu_1 &= \min(M_1^c, q_{221} + q_{212}), \\
\mu_2 &= \min(M_2^c, q_{221} + q_{122}), \\
\mu_3 &= \min(M_3^c, q_{212} + q_{122}), \\
\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) &= \min(M_1^c + q_{122}, M_2^c + q_{212}, M_3^c + q_{122})
\end{aligned}$$

(Étape 3) Si $\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221} \geq q_{221} + q_{212} + q_{122}$ alors

$$\begin{aligned}
\pi_{221} &= q_{221}, \\
\pi_{212} &= q_{212}, \\
\pi_{122} &= q_{122}, \\
\mu_1 &= q_{221} + q_{212}, \\
\mu_2 &= q_{221} + q_{122}, \\
\mu_3 &= q_{212} + q_{122},
\end{aligned}$$

(Étape 4) Si $\frac{\mu_1 + \mu_2 + \mu_3}{2} < \min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221})$ alors

$$\sup\{D'_0(u_1, u_2, u_3) : D'_0 \in \mathcal{D}\} = q_{111} + \pi_{211} + \pi_{121} + \pi_{112} + \frac{\mu_1 + \mu_2 + \mu_3}{2}.$$

En effet, nous avons

$$\begin{aligned}
q'_{221} + q'_{212} + q'_{122} &= \frac{q'_{221} + q'_{212}}{2} + \frac{q'_{212} + q'_{122}}{2} + \frac{q'_{221} + q'_{122}}{2} \\
&\leq \frac{\mu_1 + \mu_2 + \mu_3}{2}.
\end{aligned}$$

En choisissant

$$\begin{aligned} q'_{221} &= \frac{\mu_1 + \mu_2 - \mu_3}{2}, \\ q'_{212} &= \frac{\mu_1 + \mu_3 - \mu_2}{2}, \\ q'_{122} &= \frac{\mu_2 + \mu_3 - \mu_1}{2}, \end{aligned}$$

on atteint la borne. Pour tout autre solution telle que $q'_{222} = S > 0$, nous avons

$$\begin{aligned} q'_{221} + q'_{212} &\leq M_1^c - S, \\ q'_{221} + q'_{122} &\leq M_2^c - S, \\ q'_{212} + q'_{122} &\leq M_3^c - S, \end{aligned}$$

ce qui entraîne

$$\begin{aligned} q'_{221} + q'_{212} + q'_{122} &\leq \frac{M_1^c + M_2^c + M_3^c}{2} - \frac{S}{2}, \\ &\leq \frac{\mu_1 + \mu_2 + \mu_3}{2} - \frac{S}{2}. \end{aligned}$$

(Étape 5) Si $\frac{\mu_1 + \mu_2 + \mu_3}{2} \geq \min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221})$ et $\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) < q'_{221} + q'_{212} + q'_{122}$ alors

$$\begin{aligned} \sup\{D'_0(u_1, u_2, u_3) : D'_0 \in \mathcal{D}\} &= q_{111} + \pi_{211} + \pi_{121} + \pi_{112} + \\ &\quad \min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}). \end{aligned}$$

En effet, supposons sans perte de généralité que

$$\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) = \mu_1 + \pi_{122}.$$

Nous avons

$$\begin{aligned} \mu_1 &= M_1^c \\ q'_{221} + q'_{212} + q'_{122} &\leq \mu_1 + \pi_{122} = M_1^c + \pi_{122}. \end{aligned}$$

La borne est atteinte pour

$$\begin{aligned} q'_{221} &= \frac{\mu_2 - \pi_{221}}{\mu_2 + \mu_3 - 2\pi_{122}} \mu_1, \\ q'_{212} &= \frac{\mu_3 - \pi_{122}}{\mu_2 + \mu_3 - 2\pi_{122}} \mu_1 \\ q'_{122} &= \pi_{122} \end{aligned}$$

Ce choix maximise $q'_{221} + q'_{212} + q'_{122}$ et donne $q'_{222} = 0$. Si on prend un autre choix avec $q'_{222} = S > 0$ alors on devra avoir

$$\begin{aligned} q'_{221} + q'_{212} &\leq M_1^c - S, \\ q'_{221} + q'_{212} + q'_{122} + S &\leq M_1^c + \pi_{122}. \end{aligned}$$

Pour satisfaire les contraintes, on pose

$$\begin{aligned}
q'_{321} &= 0, \\
q'_{231} &= \min(M_2^c - q'_{221}, q_{221} - q'_{221}), \\
q'_{312} &= 0, \\
q'_{213} &= \min(M_3^c - q'_{212}, q_{212} - q'_{212}), \\
q'_{122} &= \min(M_2^c - \min(q'_{221}, M_2^c), M_3^c - \min(q'_{212}, M_3^c), q_{122}) \\
q'_{123} &= \min(M_2^c - \min(q'_{221}, M_2^c) - q'_{122}, q_{122} - q'_{122}) \\
q'_{132} &= \min(M_3^c - \min(q'_{212}, M_3^c) - q'_{122}, q_{122} - q'_{122}) \\
q'_{222} &= 0, \\
q'_{232} &= 0, \\
q'_{321} &= 0, \\
q'_{223} &= 0, \\
q'_{233} &= 0, \\
q'_{322} &= \min(M_2^c - \min(M_2^c, q_{212} + q_{221}), M_3^c - \min(M_3^c, q_{212} + q_{221}), q_{222}), \\
q'_{323} &= \min(M_2^c - \min(M_2^c, q_{212} + q_{221}) - q'_{322}, q_{222} - q'_{322}), \\
q'_{332} &= \min(M_3^c - \min(M_3^c, q_{212} + q_{122}) - q'_{322}, q_{222} - q'_{322}),
\end{aligned}$$

(Étape 6) Si $\min(\mu_1 + \pi_{122}, \mu_2 + \pi_{212}, \mu_3 + \pi_{221}) \geq q'_{221} + q'_{212} + q'_{122}$ alors

$$\begin{aligned}
\sup\{D'_0(u_1, u_2, u_3) : D'_0 \in \mathcal{D}\} &= q_{111} + \pi_{211} + \pi_{121} + \pi_{112} + \\
&\quad q_{122} + q_{212} + q_{221} + \pi_{222}.
\end{aligned}$$

Le choix optimal est

$$\begin{aligned}
q'_{221} &= \pi_{221}, \\
q'_{212} &= \pi_{212}, \\
q'_{122} &= \pi_{122}, \\
q'_{222} &= \pi_{222}.
\end{aligned}$$

Si on choisit $q'_{222} < \pi_{222}$ alors

$$\begin{aligned}
q'_{221} + q'_{212} + q'_{122} &\leq q_{221} + q_{212} + q_{122}, \\
q'_{221} + q'_{212} + q'_{122} + q'_{222} &< q_{122} + q_{212} + q_{221} + \pi_{222}.
\end{aligned}$$

Si on choisit $q'_{222} > \pi_{222}$ et les contraintes sur les marges satisfaites alors $q_{222} < \pi_{222}$. Supposons sans perte de généralité que $\pi_{222} = M_1^c - (q_{221} - q_{212})$.

Dans ce cas,

$$\begin{aligned}
q'_{221} + q'_{212} &\leq \pi_{221} + \pi_{212} - (q_{221} - q_{212}), \\
q'_{122} &\leq q_{122}, \\
q'_{221} + q'_{212} + q'_{122} + q'_{222} &\leq \pi_{221} + \pi_{212} + \pi_{122} + \pi_{222},
\end{aligned}$$

Si $\pi_{222} = q_{222}$ posons $q'_{222} = q_{222}$; sinon sans perte de généralité, posons

$$\begin{aligned}\pi_{222} &= M_1^c - (q_{221} + q_{212}), \\ q'_{222} &= \pi_{222}, \\ q'_{223} &= q'_{232} = q'_{233} = 0, \\ q'_{322} &= \min(M_2^c - (q_{221} + q_{122}), M_3^c - (q_{212} + q_{122}), q_{222}), \\ q'_{323} &= \min(M_2^c - (q_{221} + q_{212}) - (q'_{222} + q'_{322}), q_{222} - (q'_{222} + q'_{322})), \\ q'_{332} &= \min(M_3^c - (q_{212} + q_{122}) - (q'_{222} + q'_{322}), q_{222} - (q'_{222} + q'_{322})).\end{aligned}$$

La borne est atteinte.

□

Exemple 4.4.1. Considérons $D : \{0, 1/2, 1\} \times \{0, 1/2, 1\} \times \{0, 1\} \rightarrow [0, 1]$ telle que $D(1/2, 1/2, 1) = 1/4$.

Soit la fonction UC donnée par

— (1) pour $(v_1, v_2, v_3) \in [0, 1/2]^2 \times [0, 1]$

$$UC(v_1, v_2, v_3) = \min(v_1, v_2, v_3, \frac{1}{4});$$

— (2) pour $(v_1, v_2, v_3) \in [0, 1/2] \times [1/2, 1] \times [0, 1]$

$$UC(v_1, v_2, v_3) = \min(\min(v_1, v_3, \frac{1}{4}) + \min(v_2 - 1/2, 1/4), \min(v_1, v_3, 1/2))$$

— (3) pour $(v_1, v_2, v_3) \in [1/2, 1] \times [0, 1/2] \times [0, 1]$

$$UC(v_1, v_2, v_3) = \min(\min(v_2, v_3, \frac{1}{4}) + \min(v_1 - 1/2, 1/4), \min(v_2, v_3, 1/2))$$

— (4) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [0, 1/4]$

$$UC(v_1, v_2, v_3) = v_3$$

— (5) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(v_3 - 1/4, 1/2) > \min(v_1 - 1/2, 1/4) + \min(v_2 - 1/2, 1/4)$$

$$\begin{aligned}UC(v_1, v_2, v_3) &= \min[\min(v_3, \frac{1}{4}) + \max(0, \min(v_1 - 1/2, 1/4)) \\ &\quad + \max(0, \min(v_2 - 1/2, 1/2)), \\ &\quad \min(v_3, \frac{1}{4}) + \max(0, \min(v_2 - 1/2, 1/4)) \\ &\quad + \max(0, \min(v_1 - 1/2, 1/2))].\end{aligned}$$

— (6) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(v_3 - 1/4, 1/2) \leq \min(v_1 - 1/2, 1/4) + \min(v_2 - 1/2, 1/4)$$

$$UC(v_1, v_2, v_3) = \min(v_1, v_2, v_3).$$

Montrons que la fonction UC est un majorant

Soit C une copule prolongement de D .

— (1) Pour $(v_1, v_2, v_3) \in [0, 1/2]^2 \times [0, 1]$

$$C(v_1, v_2, v_3) \leq \min(v_1, v_2, v_3, \frac{1}{4}) = UC(v_1, v_2, v_3);$$

— (2) pour $(v_1, v_2, v_3) \in [0, 1/2] \times [1/2, 1] \times [0, 1]$

$$\begin{aligned} C(v_1, v_2, v_3) &\leq \min(V_C([0, v_1] \times [0, 1/2] \times [0, v_3]) + \\ &\quad V_C([0, v_1] \times (1/2, v_2] \times [0, v_3]), V_C([0, v_1] \times [0, 1] \times [0, v_3])) \\ &\leq UC(v_1, v_2, v_3) \end{aligned}$$

— (3) pour $(v_1, v_2, v_3) \in [1/2, 1] \times [0, 1/2] \times [0, 1]$, par symétrie pour le cas $(v_1, v_2, v_3) \in [0, 1/2] \times [1/2, 1] \times [0, 1]$

$$C(v_1, v_2, v_3) \leq UC(v_1, v_2, v_3)$$

— (4) $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [0, 1/4]$

$$C(v_1, v_2, v_3) \leq v_3 = UC(v_1, v_2, v_3)$$

— (5) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(v_3 - 1/4, 1/2) > \min(v_1 - 1/2, 1/4) + \min(v_2 - 1/2, 1/4)$$

Pour toute copule C , prolongement de D , nous avons

$$\begin{aligned} C(v_1, v_2, v_3) &\leq \min[V_C([0, 1/2] \times [0, 1/2] \times [0, v_3]) + V_C([1/2, v_1] \times [0, 1/2] \times [0, v_3]) + \\ &\quad V_C([0, v_1] \times [1/2, v_2] \times [0, v_3]), V_C([0, 1/2] \times [0, 1/2] \times [0, v_3]) + \\ &\quad V_C([0, 1/2] \times [1/2, v_2] \times [0, v_3]) + V_C([1/2, v_1] \times [0, v_2] \times [0, v_3])] \\ &\leq \min[\min(v_3, \frac{1}{4}) + \max(0, \min(v_1 - 1/2, 1/4)) + \\ &\quad \max(0, \min(v_2 - 1/2, 1/4)), \min(v_3, \frac{1}{4}) + \\ &\quad \max(0, \min(v_2 - 1/2, 1/4)) + \max(0, \min(v_1 - 1/2, 1/4))], \\ &\leq UC(v_1, v_2, v_3). \end{aligned}$$

— (6) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(v_3 - 1/4, 1/2) \leq \min(v_1 - 1/2, 1/4) + \min(v_2 - 1/2, 1/4)$$

Pour toute copule C , prolongement de D , nous avons

$$\begin{aligned} C(v_1, v_2, v_3) &\leq \min(v_1, v_2, v_3), \\ &\leq UC(v_1, v_2, v_3). \end{aligned}$$

Montrons que la fonction UC est la borne supérieure

Considérons les prolongements $C_1, C_2 : [0, 1]^3 \longrightarrow [0, 1]$, tels que pour $(v_1, v_2, v_3) \in [0, 1]^3$,

$$\begin{aligned} C_1(v_1, v_2, v_3) &= \max(0, \min(v_1, v_2, v_3, 1/4)) \\ &+ \max(0, \min(v_1 - 1/4, v_2 - 1/2, v_3 - 1/2, 1/4)) \\ &+ \max(0, \min(v_1 - 1/2, v_2 - 1/4, v_3 - 1/4, 1/4)) \\ &+ \max(0, \min(v_1 - 3/4, v_2 - 3/4, v_3 - 3/4, 1/4)); \end{aligned}$$

$$\begin{aligned} C_2(v_1, v_2, v_3) &= \max(0, \min(v_1, v_2, v_3, 1/4)) \\ &+ \max(0, \min(v_1 - 1/4, v_2 - 1/2, v_3 - 1/4, 1/4)) \\ &+ \max(0, \min(v_1 - 1/2, v_2 - 1/4, v_3 - 1/2, 1/4)) + \\ &+ \max(0, \min(v_1 - 3/4, v_2 - 3/4, v_3 - 3/4, 1/4)); \end{aligned}$$

pour (u_1, u_2, u_3) tel que $(u_1, u_2, u_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(u_3 - 1/4, 1/2) > \min(u_1 - 1/2, 1/4) + \min(u_2 - 1/2, 1/4)$$

,

$$\begin{aligned} C_3(v_1, v_2, v_3) &= \max(0, \min(v_1, v_2, v_3, 1/4)) \\ &+ \max(0, \min(v_1 - 1/4, v_2 - 1/2, v_3 - 1/4, \min(u_2 - 1/2, 1/4))) \\ &+ \max(0, \min(v_1 - 1/4 - \min(u_2 - 1/2, 1/4), \\ &\quad v_2 - 1/2 - \min(u_2 - 1/2, 1/4), \\ &\quad v_3 - 1/2 - \min(u_2 - 1/2, 1/4), 1/4 - \min(u_2 - 1/2, 1/4))) \\ &+ \max(0, \min(v_1 - 1/2, v_2 - 1/4, v_3 - 1/4 - \min(u_2 - 1/2, 1/4), \\ &\quad \min(u_1 - 1/2, 1/4))) \\ &+ \max(0, \min(v_1 - 1/2 - \min(u_1 - 1/2, 1/4), \\ &\quad v_2 - 1/4 - \min(u_1 - 1/2, 1/4), \\ &\quad v_3 - 1/4 - \min(u_2 - 1/2, 1/4) - \min(u_1 - 1/2, 1/4), \\ &\quad 1/4 - \min(u_1 - 1/2, 1/4))) \\ &+ \max(0, \min(v_1 - 3/4, v_2 - 3/4, v_3 - 3/4, 1/4)). \end{aligned}$$

On peut remarquer que

$$C_3(u_1, u_2, u_3) = 1/4 + \min(u_2 - 1/2, 1/4) + \min(u_1 - 1/2, 1/4)$$

du fait qu'on ne peut avoir à la fois $u_1 > 3/4$ et $u_2 > 3/4$.

Nous montrerons que C_1, C_2, C_3 atteignent la borne supérieure des prolongements de D sur des régions dont la réunion donne $[0, 1]^3$.

— (1) Sur $[0, 1/2]^2 \times [0, 1]$, nous avons

$$\begin{aligned} C_1(v_1, v_2, v_3) &= \max(0, \min(v_1, v_2, v_3, 1/4)) \\ &= UC(v_1, v_2, v_3); \end{aligned}$$

— (2) Sur $[1/2, 1] \times [0, 1/2] \times [0, 1]$, nous avons

$$\begin{aligned} C_1(v_1, v_2, v_3) &= \max(0, \min(v_2, v_3, 1/4)) + \\ &\quad \max(0, \min(v_1 - 1/2, v_2 - 1/4, v_3 - 1/4, 1/4)) \\ &= UC(v_1, v_2, v_3); \end{aligned}$$

pour $(v_1, v_2, v_3) \in [1/2, 1] \times [0, 1/2] \times [0, 1]$.

— (3) Sur $[0, 1/2] \times [1/2, 1] \times [0, 1]$

$$\begin{aligned} C_2(v_1, v_2, v_3) &= \max(0, \min(v_1, v_3, 1/4)) + \max(0, \min(v_1 - 1/4, v_2 - 1/2, v_3 - 1/4, 1/4)) \\ &= UC(v_1, v_2, v_3), \end{aligned}$$

pour $(v_1, v_2, v_3) \in [0, 1/2] \times [1/2, 1] \times [0, 1]$.

— (4) Sur $[1/2, 1]^2 \times [0, 1/4]$, nous avons

$$C_1(v_1, v_2, v_3) = v_3 = UC(v_1, v_2, v_3).$$

— (5) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(v_3 - 1/4, 1/2) > \min(v_1 - 1/2, 1/4) + \min(v_2 - 1/2, 1/4);$$

soit (u_1, u_2, u_3) tel que $(u_1, u_2, u_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(u_3 - 1/4, 1/2) > \min(u_1 - 1/2, 1/4) + \min(u_2 - 1/2, 1/4);$$

nous avons

$$C_3(u_1, u_2, u_3) = 1/4 + \min(u_2 - 1/2, 1/4) + \min(u_1 - 1/2, 1/4) = UC(u_1, u_2, u_3)$$

— (6) pour $(v_1, v_2, v_3) \in [1/2, 1]^2 \times [1/4, 1]$ et

$$\min(v_3 - 1/4, 1/2) \leq \min(v_1 - 1/2, 1/4) + \min(v_2 - 1/2, 1/4)$$

— (6-a) si $1/4 \leq v_3 \leq 3/4$, nous avons

$$\begin{aligned} C_1(v_1, v_2, v_3) &= \max(0, \min(v_3, 1/4)) + \max(0, \min(v_3 - 1/2, 1/4)) + \\ &\quad \max(0, \min(v_3 - 1/4, 1/4)), \\ &= v_3 = UC(v_1, v_2, v_3) \end{aligned}$$

— (6-b) si $v_3 \geq 3/4$, d'après les conditions posées, nous devrions avoir

$$v_1 \geq 3/4, v_2 \geq 3/4 \text{ et}$$

$$\begin{aligned} C_1(v_1, v_2, v_3) &= 3/4 + \max(0, \min(v_1 - 3/4, v_2 - 3/4, v_3 - 3/4, 1/4)), \\ &= \min(v_1, v_2, v_3) = UC(v_1, v_2, v_3) \end{aligned}$$

4.5. CONCLUSION

Nous nous sommes intéressés aux prolongements d'une sous-copule à domaine en dimension $d = 3$. La borne supérieure n'est pas nécessairement une copule contrairement au cas de la dimension 2. Nous donnons la formule de la borne supérieure. Celle de la borne inférieure et la généralisation au cas infini feront l'objet d'un travail ultérieur.

Chapitre 5

CONCLUSION

Cette thèse s'intéresse aux prolongements de sous-copules. L'une des applications concerne l'estimation des copules à l'aide des prolongements de type copule empirique ou copule empirique lissée. Cette méthode permet d'obtenir des estimateurs bona fide de la copule. Il y a un manque de résultats en ce qui concerne la convergence uniforme sur $[0, 1]^d$ et la convergence uniforme locale (sur un voisinage autour d'un point) de tels estimateurs. Le chapitre 2 donne le résultat local en s'appuyant sur le lissage spline de la copule empirique, tandis que le résultat sur $[0, 1]^d$ est donné par l'estimateur copule de Bernstein empirique. C'est un premier pas vers la construction d'un estimateur multirésolution par insertion de nœuds à partir de la copule de Bernstein empirique.

Le chapitre 3 s'est attelé à caractériser l'ensemble des copules qui prolongent une sous-copule donnée en dimension 2 et donne la borne supérieure et la borne inférieure de cet ensemble. La borne supérieure et la borne inférieure de tels prolongements se trouvent être elles aussi des prolongements de la sous-copule.

Le chapitre 4 essaie de déterminer la borne supérieure des prolongements d'une sous-copule de domaine fini en dimension 3. Cette borne supérieure se trouve être une quasi-copule. Un exemple de la construction d'une telle borne est donné à la fin du chapitre. Une des applications du présent travail serait l'étude du problème d'incompatibilité : la difficulté de construire des fonctions de répartition ayant des marges multivariées fixées. Une première étape serait de construire la borne supérieure et la borne inférieure point par point de l'ensemble de ces fonctions de répartition lorsque cet ensemble est non vide. Cette thèse pourrait fournir les matériaux pour une telle construction.

Annexe A

UNE APPROCHE PROBABILISTE DES SPLINES

A.1. CONSTRUCTION DE SPLINES MONOTONES

Considérons un ensemble de nœuds

$$0 < x_1 \leq x_2 \leq \dots \leq x_m < 1, \text{ avec } m > 0.$$

supposons que y_1, \dots, y_{m^*} sont les valeurs distinctes de x_1, \dots, x_m , avec y_1 de multiplicité k_1, \dots, y_{m^*} de multiplicité $k_{m^*} \leq k$. Pour $0 < k_0$ et $0 < k_{m^*+1}$, posons

$$x_j = \begin{cases} 0 & \text{pour } j = 0, -1, \dots, -k_0 + 1 \\ 1 & \text{pour } j = m + 1, m + 2, \dots, m + k_{m^*+1}, \end{cases}$$

$I_{ij} = [x_{j-i}, x_j)$, $j = 1, \dots, m + i$, $i = 1, \dots, k$ et k est un réel. Pour simplifier la présentation, nous prendrons $k_0 = k_{m^*+1} = k$.

Nous construisons des fonctions de répartition $G_{j,i}$, $j = 1, \dots, m + i$, $i = 1, \dots, k$ en deux étapes. Considérons aussi les fonctions de répartition $H_{j,i}$, $j = 1, \dots, m + i$, $i = 1, \dots, k$ définies par les formules

$$H_{j,i}(x) = \begin{cases} 0 & \text{si } x < x_{j-i}, \\ \frac{x - x_{j-i}}{x_j - x_{j-i}} & \text{si } x \in I_{ij}, \ x_{j-i} < x_j, \\ I(x \geq x_j) & \text{si } x \in I_{ij}, \ x_{j-i} = x_j. \\ 1 & \text{si } x > x_j. \end{cases}$$

Les splines polynomiales $\{G_{j,i}, j = 1, \dots, m + i, i = 1, \dots, k_i\}$ sont construites selon les formules de récurrence :

$$G_{j,i+1} = \begin{cases} H_{j,i+1} + (1 - H_{j,i+1})G_{j,i} & \text{si } j = 1, \\ H_{j,i+1}G_{j-1,i} + (1 - H_{j,i+1})G_{j,i} & \text{si } j = 2, \dots, m + i \\ H_{j,i+1}G_{j-1,i} & \text{si } j = m + i + 1, \end{cases} \quad (\text{A.1.1})$$

A-ii

$j = 1, \dots, m + i, i = 1, \dots, k$, et $G_{j,0} = I(x \geq x_j)$, $j = 1, \dots, m$. Si $m = 0$ et que

$$0 = x_{1-k} = \dots = x_0 < 1 = x_1 = x_2 = \dots = x_k,$$

les relations ressemblent aux relations de récurrence définissant les polynômes de Bernstein :

$$G_{j,i+1} = xG_{j-1,i} + (1-x)G_{j,i} \text{ si } j = 2, \dots, i.$$

Ces relations peuvent être mises en parallèles avec les formules de récurrence [de Boor, 1972] définissant les B-splines normalisées ([Schoenberg, 1946]) :

$$N_{j-i-1,i+1}(x) = \frac{x - x_{j-i-1}}{x_{j-1} - x_{j-i-1}} N_{j-i-1,i}(x) + \frac{x_j - x}{x_j - x_{j-i}} N_{j-i,i}(x),$$

$x \in [x_{j-i-1}, x_j)$, et $N_{j,1} = I_{[x_j, x_{j+1})}$ avec $j = 1, \dots, m + i, i = 0, \dots, k$.

Lemme A.1.1. *Nous avons*

$$G_{j,i} - G_{j+1,i} = N_{j-i,i+1} \quad (\text{A.1.2})$$

$$G_{j,i} = \sum_{l=j}^{m+i} N_{l-i,i+1}, \quad (\text{A.1.3})$$

$$g_{j,i} = \frac{i}{x_j - x_{j-i}} N_{j-i,i} = M_{j-i,i},$$

avec $g_{i,j} = G'_{i,j}$, $M_{j-i,i}$ la B-spline non normalisée ou encore M-spline associée à $N_{j-i,i}$, $j = 0, \dots, m + i, i = 1, \dots, k$.

DÉMONSTRATION. Nous allons tout d'abord montrer que $G_{m+i,i} = N_{m,i+1}$. D'après les relations de définition (A.1.4)

$$G_{m+i,i} = H_{m+i,i} G_{m+i-1,i-1}, \quad (\text{A.1.4})$$

avec $G_{m,0} = I(x_m \leq x < x_{m+1}) = I(x_m \leq x < 1)$. Les splines $N_{m,i+2}, i = 0, \dots, k$ sont définies par les relations de récurrence :

$$N_{m,i+1}(x) = \frac{x - x_m}{x_{m+i} - x_m} N_{m,i}(x),$$

avec $N_{m,1} = I(x_m \leq x < x_{m+1}) = I(x_m \leq x < 1)$. On peut déduire que

$$G_{m+i,i} = N_{m,i+1},$$

$i = 0, \dots, k$.

De façon générale, nous allons montrer que $G_{j,i} - G_{j+1,i} = N_{j-i,i+1}$. Pour $x \in [x_{j-i}, x_j)$, en utilisant les équations de récurrence (A.1.4), nous pouvons

écrire

$$\begin{aligned}
G_{j,i+1}(x) - G_{j+1,i+1}(x) &= \left(\frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} G_{j-1,i}(x) + \frac{x_j-x}{x_j-x_{j-i-1}} G_{j,i}(x) \right) \\
&\quad - \left(\frac{x-x_{j-i}}{x_{j+1}-x_{j-i}} G_{j,i}(x) + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} G_{j+1,i}(x) \right), \\
&= \frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} (G_{j-1,i} - G_{j,i})(x) + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} (G_{j,i} - G_{j+1,i})(x) \\
&\quad + \left(\frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} - \frac{x-x_{j-i}}{x_{j+1}-x_{j-i}} \right) G_{j,i}(x) + \left(\frac{x_j-x}{x_j-x_{j-i-1}} - \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} \right) G_{j+1,i}(x), \\
&= \frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} (G_{j-1,i} - G_{j,i})(x) + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} (G_{j,i} - G_{j+1,i})(x).
\end{aligned}$$

D'autre part, si $x \in [x_{j-i-1}, x_{j-i})$, $G_{j+1,i+1}(x) = G_{j,i}(x) = G_{j+1,i}(x) = 0$ et nous avons

$$\begin{aligned}
(G_{j,i+1} - G_{j+1,i+1})(x) &= \frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} G_{j-1,i}(x), \\
&= \frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} (G_{j-1,i} - G_{j,i})(x) + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} (G_{j,i} - G_{j+1,i})(x).
\end{aligned}$$

Si $x \in [x_j, x_{j+1})$, $G_{j,i+1}(x) = G_{j,i}(x) = G_{j-1,i}(x) = 1$, nous avons

$$\begin{aligned}
(G_{j,i+1} - G_{j+1,i+1})(x) &= \left(\frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} + \frac{x_j-x}{x_j-x_{j-i-1}} \right) - \left(\frac{x-x_{j-i}}{x_{j+1}-x_{j-i}} + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} G_{j+1,i}(x) \right), \\
&= \frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} (G_{j-1,i} - G_{j,i})(x) + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} (G_{j,i} - G_{j+1,i})(x).
\end{aligned}$$

En résumé, pour $x \in [x_{j-i-1}, x_{j+1})$,

$$(G_{j,i+1} - G_{j+1,i+1})(x) = \frac{x-x_{j-i-1}}{x_j-x_{j-i-1}} (G_{j-1,i} - G_{j,i})(x) + \frac{x_{j+1}-x}{x_{j+1}-x_{j-i}} (G_{j,i} - G_{j+1,i})(x).$$

De plus $G_{j,0} - G_{j+1,0} = N_{j,1} = I_{[x_j, x_{j+1})}$, $j = 1, \dots, m+i$, $i = 0, \dots, k$. Les splines $\{G_{j,i} - G_{j+1,i}, j = 0, \dots, m+i, i = 0, \dots, k\}$ vérifient les relations de définition des b-splines normalisées et leur sont égales pour $i = 0$. Par induction, ce sont des B-splines normalisés. La deuxième assertion vient de la somme télescopique.

Nous allons prouver la dernière assertion par récurrence sur j . Tout d'abord, montrons qu'elle est vraie pour $j = m+i$. En dérivant l'équation

$$G_{j,i} - G_{j+1,i} = N_{j-i,i+1},$$

nous avons

$$g_{j,i} - g_{j+1,i} = N'_{j-i,i+1} = \frac{i}{x_j - x_{j-i}} N_{j-i,i} - \frac{i}{x_{j+1} - x_{j-i+1}} N_{j-i+1,i},$$

car d'après [Prochazkova, 2005]

$$N'_{l-i,i+1} = \frac{i}{x_l - x_{l-i}} N_{l-i,i} - \frac{i}{x_{l+1} - x_{l-i+1}} N_{l-i+1,i}.$$

Ainsi,

$$g_{m+i,i} - g_{m+i+1,i} = (i/(x_{m+i} - x_m)) N_{m,i} - (i/(x_{m+i+1} - x_{m+1})) N_{m+1,i},$$

A-iv

et puisque $g_{m+i+1,i} = 0 = N_{m+1,i}$, nous avons

$$g_{m+i,i} = (i/(x_{m+i} - x_m))N_{m,i}.$$

Supposons maintenant que pour j fixé,

$$g_{j,i} = \frac{i}{x_j - x_{j-i}} N_{j-i,i},$$

et montrons qu'alors

$$g_{j-1,i} = \frac{i}{x_{j-1} - x_{j-i-1}} N_{j-i-1,i},$$

À partir des égalités

$$g_{j,i} - g_{j+1,i} = N'_{j-i,i+1} = \frac{i}{x_j - x_{j-i}} N_{j-i,i} - \frac{i}{x_{j+1} - x_{j-i+1}} N_{j-i+1,i},$$

$$g_{j-1,i} - g_{j,i} = N'_{j-i-1,i+1} = \frac{i}{x_{j-1} - x_{j-i-1}} N_{j-i-1,i} - \frac{i}{x_j - x_{j-i}} N_{j-i,i},$$

et de l'hypothèse de récurrence, le résultat se déduit. \square

Dès lors, si $m = 0$ et que

$$0 = x_{1-k} = \dots = x_0 < 1 = x_1 = x_2 = \dots = x_k,$$

les fonctions $\{g_{j,i}, j = 1, \dots, i-1\}$ sont des polynômes de Bernstein et les fonctions $\{G_{j,i}, j = 1, \dots, i-1\}$ sont des primitives de polynômes de Bernstein s'annulant en 0 et valant 1 au point 1. De façon générale, lorsqu'elles sont continues, les fonctions $\{G_{j,i}, j = 0, \dots, m+i+1, i = 1, \dots, k\}$ sont des I-splines.

Corollaire A.1.1. *Les splines $\{G_{j,i}, j = 0, \dots, m+i+1, i = 1, \dots, k\}$, lorsqu'elles sont continues, sont des I-splines et $\{g_{j,i}, j = 0, \dots, m+i+1, i = 1, \dots, k\}$ sont des b-splines non normalisés ou encore M-splines.*

L'intérêt de construire les splines monotones par les relations de récurrence tient à leur mise en oeuvre. En effet, les I-splines sont obtenus par intégration des M-splines et lorsque les nœuds sont simples, s'écrivent comme une somme de b-splines normalisées. La formule (A.1.3) généralise cette écriture aux I-splines avec nœuds multiples. Les formules (A.1.4) et (A.1.3) fournissent des algorithmes de calcul simple des I-splines. En outre, ces splines monotones peuvent présenter des discontinuités, ce qui n'est pas le cas des I-splines. Elles sont donc une généralisation possible des I-splines.

Les splines ainsi construites serviront à approcher des fonctions monotones notamment les fonctions de répartition.

A.2. APPROXIMATION À L'AIDE DES SPLINES MONOTONES

Les splines monotones décrites à la section précédente permettent de reconstruire parfaitement les fonctions de répartition qui sont des polynômes. Le lemme suivant donne le cas particulier de la fonction de répartition uniforme.

Lemme A.2.1. *Les fonctions de répartition $\{G_{j,k}\}_{j=1}^{m+k}$, satisfont la condition :*

$$\sum_{j=1}^{m+k} \frac{x_j - x_{j-k}}{k} G_{j,k}(x) = x, \quad x \in [0, 1]$$

DÉMONSTRATION. C'est une conséquence de l'identité de Marsden [Marsden, 1968] pour le polynôme de degré un. \square

L'approximation d'une fonction de répartition F à l'ordre k dépend de poids $\lambda_{j,k}$, $j = 1, \dots, m + k_{m+1}$ et est donnée par

$$G_k = \sum_{j=1}^{m+k_{m+1}} \lambda_{j,k} G_{j,k}.$$

Lorsque les nœuds intérieurs sont de multiplicité un ($k_1 = \dots = k_m$) et $k_0 = k_{m+1} = k$, Schoenberg [Schoenberg, 1967] a défini une méthode d'approximation par splines d'une fonction F définie sur $[0, 1]$ par la formule :

$$Q_\pi F(x) = \sum_{j=1}^{m+k} F(\xi_j) N_{j-k-1, k+1}(x),$$

avec ξ_j vérifiant

$$\sum_{j=1}^{m+k} \xi_{j-k} N_{j-k, k}(x) = x, \quad x \in [0, 1],$$

$$\xi_j = \frac{1}{k} \sum_{l=1}^k x_{ij+l},$$

$j = 1 - k, \dots, m$ et π la partition de l'axe engendrée par les nœuds.

Une écriture équivalente avec les splines monotones définies ci-dessus (en utilisant l'équation (A.1.2)) est :

$$Q_\pi F(x) = \sum_{j=1}^{m+k} (F(\xi_{j-k+1}) - F(\xi_{j-k})) G_{j,k}(x). \quad (\text{A.2.1})$$

Ce type d'approximation inclut l'approximation par polynômes de Bernstein comme cas particulier. [Marsden, 1972] a montré que pour cette méthode, lorsque

les nœuds intérieurs sont de multiplicité un, l'erreur en norme uniforme est majorée de la façon suivante :

$$\|Q_\pi F - F\|_\infty < \omega(F, h)(1 + \min\{\frac{1}{h\sqrt{(2k-2)}}, \frac{\sqrt{\frac{k}{12}}|\pi|}{h}\});$$

ou si $F \in C[0, 1]$

$$\|Q_\pi F - F\|_\infty < \omega_2(F, h)(1 + \frac{1}{h^2} \min\{\frac{1}{2k}, \frac{(k+1)|\pi|^2}{12}\})$$

avec $|\pi| = \max_j (x_{j+1} - x_j)$ le pas maximal de la partition π en gendrée par les nœuds $0 = x_0 \leq x_1 \leq \dots \leq x_m \leq x_{m+1} = 1$.

Le terme $\min\{\frac{1}{|\pi|\sqrt{(2k-2)}}, \sqrt{\frac{k}{12}}\}$ indique que le comportement asymptotique est proche de celui des splines polynomiaux pour $|\pi|$ petit. Lorsque $|\pi|$ est grand, le comportement asymptotique est proche de celui des polynômes de Bernstein [de Leeuw, 1959]. [Beutel et coll., 2002] donne l'erreur d'approximation de la dérivée $(Q_\pi F)'$:

$$\|(Q_\pi F - F)'\|_\infty < \frac{|\pi_k|}{kh} \omega_1(F', h) + (1 + \frac{1}{h^2}(1 + \sqrt{\frac{k}{12}})^2 \frac{|\pi_k|^2}{k^2}) \omega_2(F', h)$$

avec $|\pi| = \max_j |x_{j-k} - x_j|$.

Annexe B

DÉMONSTRATION DU LEMME 3.2.2

DÉMONSTRATION. Puisque \overline{D} est une sous-copule sur $\overline{S}_1 \times \overline{S}_2$, nous obtenons que V_D est positive sur \mathcal{A} et $V_D([0, 1]^2) = 1$. Nous avons besoin de montrer que V_D est σ -additive sur \mathcal{A} . Premièrement, montrons que V_D est additive sur \mathcal{A} . Soient A_1, A_2 deux ensembles disjoints de \mathcal{A} , avec $A_k = B_k \times C$, $k = 1, 2$ et $A_1 \cup A_2 \in \mathcal{A}$. Il est facile de vérifier que $V_D(A_1 \cup A_2) = V_D(A_1) + V_D(A_2)$. De façon similaire, si $A_k = B \times C_k$, $k = 1, 2$ et $A_1 \cup A_2 \in \mathcal{A}$ alors $V_D(A_1 \cup A_2) = V_D(A_1) + V_D(A_2)$. Soit $A \in \mathcal{A}$. Considérons une partition de A de la forme $A = \bigcup_{i=1}^n A_i$, $A_i \in \mathcal{A}$, $A_i = B_i \times C$, $i = 1, \dots, n$. Par induction, nous obtenons que $V_D(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n V_D(A_i)$. De la même façon, si $A = \bigcup_{i=1}^n A_i$, $A_i \in \mathcal{A}$, $A_i = B \times C_i$, $i = 1, \dots, n$, alors $V_D(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n V_D(A_i)$. Soit $A \in \mathcal{A}$. Considérons une partition de A de la forme $A = \bigcup_{i=1}^{n_1} \bigcup_{j=1}^{n_2} A_{ij}$, $A_{ij} \in \mathcal{A}$, $A_{ij} = B_i \times C_j$, $i = 1, \dots, n_1$, $j = 1, \dots, n_2$. Nous obtenons que $V_D(\bigcup_{i=1}^{n_1} \bigcup_{j=1}^{n_2} A_{ij}) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} V_D(A_{ij})$. Plus généralement, soit $A \in \mathcal{A}$, considérons une partition de A de la forme $A = \bigcup_{i=1}^{n_1} \bigcup_{j=1}^{n_2} A_{ij}$, $A_{ij} \in \mathcal{A}$, $A_{ij} = B_i \times B_j$, avec

$$A_k = \bigcup_{\{i: B_i \times [0, 1] \cap A_k \neq \emptyset\}} \bigcup_{\{j: [0, 1] \times C_j \cap A_k \neq \emptyset\}} B_i \times C_j,$$

$k = 1, \dots, n$. Dans ce cas, nous avons

$$\begin{aligned} V_D(\bigcup_{i=1}^{n_1} \bigcup_{j=1}^{n_2} A_{ij}) &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} V_D(A_{ij}) \\ &= \sum_{k=1}^n \sum_{\{(i,j): B_i \times C_j \cap A_k \neq \emptyset\}} V_D(A_{ij}) \\ &= \sum_{k=1}^n V_D\left(\bigcup_{\{(i,j): B_i \times C_j \cap A_k \neq \emptyset\}} A_{ij}\right) \\ &= \sum_{k=1}^n V_D(A_k). \end{aligned}$$

Ceci montre que V_D est additive sur \mathcal{A} . Montrons que V_D est σ -additive sur \mathcal{A} . Supposons que $A = (a, b] \times (c, d]$, $A = (a_i, b_i] \times (c_i, d_i]$, $i \geq 1$. Soit

$$B_n = (a_n - \epsilon 2^{-(n+2)}, b_n + \epsilon 2^{-(n+2)}) \times (c_n - \epsilon 2^{-(n+2)}, d_n + \epsilon 2^{-(n+2)}) \cap A.$$

B-ii

Considérons \mathring{B}_n , $n \geq 1$. Les ensembles \mathring{B}_n , $n \geq 1$ forment un recouvrement de \overline{A} . Puisque \overline{A} est fermé, il existe $N \in \mathbb{N}$ tel que $\mathring{B}_1, \mathring{B}_2, \dots, \mathring{B}_N$ est un recouvrement de \overline{A} . À fortiori, $\mathring{B}_1, \mathring{B}_2, \dots, \mathring{B}_N$ est un recouvrement de A . Nous obtenons que

$$\begin{aligned} V_D(A) &= V_D\left(\bigcup_{n=1}^N \mathring{B}_n\right) \leq \sum_{n=1}^N V_D(\mathring{B}_n) \\ &\leq \sum_{n=1}^N (V_D(A_n) + \epsilon 2^{-n}) \leq \sum_{n=1}^N V_D(A_n) + \epsilon. \end{aligned}$$

Puisque $V_D\left(\bigcup_{n=1}^N A_n\right) \leq V_D(A)$ pour tout $N \geq 1$, nous avons le résultat. Finalement, il peut être prouvé que ce résultat est aussi vrai pour $A \in \mathcal{A}$ et $A = [0, b] \times (c, d]$ ou $(a, b] \times (0, d]$ ou $[0, b] \times [0, d]$. \square

Bibliographie

- Adell, J. et Sanguesa, C. (2005). Approximation by B-spline convolution operators. A probabilistic approach. *Journal of Computational and Applied Mathematics*, **174**, 79 – 99.
- Aliprantis, D. (1998). *Principles of real analysis*. Academic Press.
- Alsina, C., Nelsen, R. B. et Schweizer, B. (1993). On the characterization of a class of binary operations on distribution functions. *Statistics and Probability Letters*, **17**, issue 2, 85–89.
- Bass, J. (1955). Sur la compatibilité des fonctions de répartition. *C. R. Acad.*, **240**, 839–841.
- Beutel, L., Gonska, H., Kacsó, D. et Tachev, G. (2002). On variation-diminishing Schoenberg operators : new quantitative statements. *Monografias de la Academia de Ciencias de Zaragoza*, **20**, 9–58.
- Bouezmarni, T., Rombouts, J. V. et Taamouti, A. (2010). Asymptotic properties of the Bernstein density copula estimator for α -mixing data. *Journal of Multivariate Analysis*, **101**, 1–10.
- Cai, B. et Meyer, R. (2011). Bayesian semiparametric modeling of survival data based on mixtures of B-spline distributions. *Computational Statistics & Data Analysis*, **55**, 1260 – 1272.
- Cameron, A., Li, T., Trivedi, P. et Zimmer, D. (2004). Modelling the differences in counted outcomes using bivariate copula models with application to mis-measured counts. *The Econometrics Journal*, **7(2)**, 566–584.
- Carley, H. (2002). Maximum and minimum extensions of finite subcopulas. *Communications in Statistics - Theory and Methods*, **31**, 2151–2166.
- Carley, H. et Taylor, M. D. (2002). A new proof of Sklar’s theorem. In *Distributions With Given Marginals and Statistical Modeling* (eds. C. Cuadras, J. Ortiana et J. Rodriguez-Lallena), 29–34. Kluwer Academic Publishers.
- Charpentier, A., Fermanian, J.-D. et Scaillet, O. (2007). The Estimation of Copulas : Theory and Practice. In *Copulas : From theory to application in finance* (ed. J. Rank), 35–62. Risk Books.

- Chen, X. et Fan, Y. (2006). Estimation of copula-based semiparametric time series models. *Journal of Econometrics*, **130**, 307 – 335.
- Dall’Aglia, G. (1972). Fréchet classes and compatibility of distribution functions. In *Symposia Mathematica, vol. 9 (Convegno di Calcolo delle Probabilità, INDAM, Rome, 1971)*. London, Academic Press.
- de Amo, E., Díaz Carrillo, M. et Fernández-Sánchez, J. (2012). Characterization of all copulas associated with non-continuous random variables. *Fuzzy Sets and Systems*, **191**, 103–112.
- de Boor, C. (1972). On calculating with B-splines. *J. Approximation Theory*, **6**, 50–62. Collection of articles dedicated to J. L. Walsh on his 75th birthday, V (Proc. Internat. Conf. Approximation Theory, Related Topics and their Applications, Univ. Maryland, College Park, Md., 1970).
- de Leeuw, K. (1959). On the degree of approximation by Bernstein polynomials. *Journal d’Analyse Mathématique*, **7**, 89–104. 10.1007/BF02787682.
- Deheuvels, P. (1978). Caractérisation complète des lois extrêmes multivariées et de la convergence des types extrêmes. *Publications de l’Institut de Statistique de l’Université de Paris*, **f.3**, 1–36.
- Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance. *Acad. Roy. Belg. Bull. Cl. Sci. (5)*, **65**, 274–292.
- Deheuvels, P. (1997). Strong laws for local quantile processes. *The Annals of Probability*, **25**, pp. 2007–2054. URL <http://www.jstor.org/stable/2959523>.
- Deheuvels, P. et Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. *The Annals of Probability*, **20**, pp. 1248–1287.
- Denuit, M. et Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis*, **93**, 40–57.
- Devroye, L. et Györfi, L. (1985). *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics : Tracts on Probability and Statistics. New York : John Wiley & Sons Inc. The L_1 view.
- Durante, F., Klement, P. et Quesada-Molina, J. (2008). Bounds for trivariate copulas with given bivariate marginals. *Journal of Inequalities and Applications*, **2008**.
- Durante, F., Saminger-Platz, S. et Sarkoci, P. (2009). Rectangular patchwork for bivariate copulas and tail dependence. *Communications in Statistics-theory and Methods*, **38**, 2515–2527.
- Einmahl, J. et Ruymgaart, F. (1987). The almost sure behaviour of the oscillation modulus of the multivariate empirical process. Open Access publications from

- Tilburg University 12-142046, Tilburg University.
- Faugeras, O. (2012). Probabilistic constructions of discrete copulas.
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, vol. 1. John Wiley & Sons : New York, 3 edn.
- Fermanian, J.-D., Radulovic, D. et Wegkamp, M. (2002). Weak convergence of empirical copula processes. Working Papers 2002-06, Centre de Recherche en Economie et Statistique.
- Fréchet, M. (1956). Sur les tableaux de corrélation dont les marges sont données. *C. R. Acad. Sci. Paris*, **242**, 2426–2428.
- Genest, C., Ghoudi, K. et Rivest, L.-P. (1995a). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, pp. 543–552.
- Genest, C., Masiello, E. et Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance : Mathematics and Economics*, **44**, 170–181.
- Genest, C., Molina, J. Q., Lallena, J. R. et Sempi, C. (1999). A characterization of quasi-copulas. *Journal of Multivariate Analysis*, **69**, 193 – 205.
- Genest, C. et Nešlehová, J. (2007). A primer on copulas for count data. *Astin Bulletin*, **37**, 475–515.
- Genest, C., Quesada-Molina, J. et Rodrigues-Lallena, J. (1995b). De l'impossibilité de construire des lois à marges multidimensionnelles données à partir de copules. *C. R. Acad. Sci. Paris*, **320**, 723–726.
- Gijbels, I. et Mielniczuk, J. (1990). Estimating the density of a copula function. *Communications in Statistics - Theory and Methods*, **19**, 445–464.
- Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24**, 726–748.
- Hoeffding, W. (1940). Massstabinvariante korrelationtheorie. *Schriften Math. Inst. Univ. Berlin*, **5**, 181–233.
- Hominal, P. et Deheuvels, P. (1979). Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Rev. Statist. Appl.*, **27**, 47–68.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London : Chapman and Hall.
- Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.*, **94**, 401–419.
- John, R. (1984). Boundary modification for kernel regression. *Communications in Statistics - Theory and Methods*, **13**, 893–900.
- Karlin, S., Micchelli, C. A. et Rinott, Y. (1986). Multivariate splines : a probabilistic perspective. *J. Multivar. Anal.*, **20**, 69–90.

- Kellerer, H. (1964). Maßtheoretische marginalprobleme. *Math. Ann.*, **153**, 168–198.
- Marsden, M. J. (1968). *An identity for spline functions with applications to variation diminishing spline approximation*. Ph.D. thesis, University of Wisconsin-Madison.
- Marsden, M. J. (1972). On uniform spline approximation. *Journal of Approximation Theory*, **6**, 249 – 253.
- McHale, I. et Scarf, P. (2007). Modelling soccer matches using bivariate discrete distributions with general dependence structure. *Statistica Neerlandica*, **61(4)**, 432–445.
- Mesfioui, M. et Quessy, J.-F. (2010). Concordance measures for multivariate non-continuous random vectors. *Journal of Multivariate Analysis*, **101**, 2398–2410.
- Mesiar, R. et Szolgay, J. (2004). W-ordinals sum of copulas and quasi-copulas. In *MAGIA 2004 Conf. Kočovce, Slovak Republic*.
- Molina, J. J. Q. et Sempi, C. (2005). Discrete quasi-copulas. *Insurance : Mathematics and Economics*, **37**, 27 – 41. Papers presented at the DeMoSTAFI Conference, Québec, 20-22 May 2004.
- Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78**, 521–530.
- Nelsen, R. (1999). *An Introduction to Copulas (Lecture Notes in Statistics)*. Springer, 1 edn.
- Nelsen, R., Quesada-Molina, J., Rodríguez-Lallena, J. et Úbeda-Flores, M. (2008). On the construction of copulas and quasi-copulas with given diagonal sections. *Insur. Math.*
- Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis*, **98**, 544 – 567.
- Niewiadomska-Bugaj, M. et Kowalczyk, T. (2005). On grade transformation and its implications for copulas. *Brazilian Journal of Probability and Statistics*, **19**, 125–137.
- Nikoloulopoulos, A. K. et Karlis, D. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, **139**, 3878–3890.
- Perron, F. et Mengersen, K. (2001). Bayesian nonparametric modeling using mixtures of triangular distributions. *Biometrics*, **57**, 518–528.
- Prochazkova, J. (2005). Derivative of B-spline function. In *25th Conference on Geometry and Graphics*.
- Rödel, E. (1987). R-estimation of normed bivariate density functions. *Statistics*, **18**, 573–585.

- Rodríguez-Lallena, J. et Úbeda Flores, M. (2004). Compatibility of three bivariate quasi-copulas : Applications to copulas. **26**, 173–180. URL http://dx.doi.org/10.1007/978-3-540-44465-7_20.
- Rodríguez-Lallena, J. et Úbeda Flores, M. (2007). Best-possible bounds on sets of multivariate distribution functions. *Communications in Statistics - Theory and Methods*, **33** :4, 805–820.
- Rüschendorf, L. (1974). On the empirical process of multivariate, dependent random variables. *J. Multivariate Anal.*, **4**, 469–478.
- Rüschendorf, L. (1976). Asymptotic distributions of multivariate rank order statistics. *Ann. Statist.*, **4**, 912–923.
- Rüschendorf, L. (2009). On the distributional transform, Sklar’s theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, **139**, issue **11**, 3921–3927.
- Sancetta, A. et Satchell, S. (2004). The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric Theory*, **20**, 535–562.
- Schoenberg, I. J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions, Part A : On the problem of smoothing of graduation, a first class of analytic approximation. *Quarterly of Applied Mathematics*, **4**, 45–88.
- Schoenberg, I. J. (1967). On spline functions. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)*, 255–291. New York : Academic Press.
- Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods*, **14**, 1123–1136.
- Schweizer, B. et Sklar, A. (1974). Operation on distribution functions not derivable from operations on random variables. *Studia Math.*, **52**, 43–52.
- Shen, X., Zhu, Y. et Song, L. (2008). Linear B-spline copulas with applications to nonparametric estimation of copulas. *Computational Statistics & Data Analysis*, **52**, 3806–3819.
- Shih, J. H. et Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, **51**, pp. 1384–1399.
- Shorack, G. R. (2000). *Probability for statisticians*. Springer.
- Sklar, A. (1996). *Distributions with fixed marginals and related topics*, chap. Random variables, distribution functions, and copulas—a personal look backward and forward, 1–14. Institute of Mathematical Statistics.

- Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris*, **8**, 229–231.
- Song, P. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, **27(2)**, 305–320.
- Stute, W. (1984). The oscillation behavior of empirical processes : The multivariate case. *The Annals of Probability*, **12**, pp. 361–379. URL <http://www.jstor.org/stable/2243477>.
- Szczesny, W. (1991). On the performance of a discriminant function. *Journal of Classification*, **8**, 201–215.
- Van Ophem, H. (1999). A general method to estimate correlated discrete random variables. *Econometric Theory*, **15(2)**, 228–237.
- Zari, T. (2010). *Contribution à l'étude du processus empirique de copule*. Ph.D. thesis, Université Paris 6.
- Zimmer, D. et Trivedi, P. (2006). Using trivariate copulas to model sample selection and treatment effects : Application to family health care demand. *Journal of Business & Statistics*, **24(1)**, 63–76.